Response to the Comment by Joakim Westerlund on "Testing for Breaks in Cointegrated Panels - with an Application to the Feldstein-Horioka Puzzle" by Francesca Di Iorio and Stefano Fachin

First of all we would like to thank Joakim Westerlund for his accurate Comment including many helpful suggestions. We will now try to answer in turn to the various questions raised, introducing for convenience each point with quotes from the relevant part of the comment (interpolated when necessary for the sake of clarity). Obviously, all the following discussion has been included in the revised version of the paper.

1 Major comments

1. "*Why even though the test takes no breaks as the null hypothesis, one still has to estimate breaks in step 2 of the bootstrap algorithm?*" To implement the bootstrap we need zero mean, stationary residuals under both H_0 and H_1. Only the residuals delivered by the more general model, allowing for a break in both constant and slope, fulfill this requirement, while those delivered by the model under no breaks do not. The latter point is easily seen:

- Data Generating Process under H_1: "break in constant":

$$
\begin{aligned}
y_t &= a_1 + a_2 d_t + b x_t + \epsilon_t, \\
d_t &= 0 \ for \ t < t_b, d_t = 1 \ else.
\end{aligned}
$$

- Model under H_0 : "no breaks": $y_t = a_1 + b x_t + e_t$

- Estimated residuals:

$$
\begin{aligned}
\widehat{e}_t &= y_t - \widehat{y}_t \\
&= (a_1 + a_2 d_t + b x_t + \epsilon_t) - (\widehat{a}_1 + \widehat{b} x_t) \\
&= (a_1 - \widehat{a}_1) + a_2 d_t + (b - \widehat{b}) x_t + \epsilon_t
\end{aligned}
$$

so that $E(\widehat{e}_t)$ has a break in t_b.

Even worse, in the case of a break in slope the residuals under H_0 will be non stationary.

On the other hand, if H_0 holds the coefficients of the model under H_1 will converge to the true parameter values.

2. *Step 3 of the bootstrap algorithm (choice of the panel summary statistic).* The mean and the median are suitable summary statistics when the aim of the panel procedure is, in Pedroni's (2004) spirit, assessing

where the mass of the distribution of the individual statistics lies. This is especially true for the median. We completely agree with Westerlund about the possible use of extremes of the individual statistics.

3. *Page 4, remark (i) (breaks estimation).* The median of the individual breakpoints will tend to fall close to $T/2$ only when H_0: "no breaks" holds (Fig. 2 in the paper). When there actually is a break, the median breakpoint will by definition identify the centre of the distribution of breakpoints, wherever that is.

4. *Section 3.1 (Monte Carlo design)* We realize that in an effort to highlight the links with the classical Gonzalo (1994) DGP we actually confused our readers. This is why we wrote the DGP in a very general form, although, as $a = 0$, the DGP actually used in the simulations is much simpler (and X is indeed not breaking). Finally, the deterministic kernel of X is governed by a parameter (θ) included in the equation of idiosyncratic noise, rather than directly in equation (3), for a better control of the panel structure. Although here homogeneity is assumed, given this structure heterogeneity would be easily introduced. This approach has been followed also in Pesaran (2007).

5. *Page 9, case 5 (Null Hypothesis of the panel test).* Here we should simply admit that we misplaced our reasoning, which applies to H_1.

6. *Page 9, paragraph 2 from the bottom (comparison between rejection rates of time series and panel tests).* The rejection rates for the time series tests are computed considering the outcomes of the individual tests on all 40 units, and are thus comparable only with the panel tests for $N = 40$. Comparisons with panel tests with smaller N are not legitimate. We could have computed the asymptotic rejection rates allowing comparisons for each sample size, but since the relative ranking of the two type of tests is obvious we preferred to keep the results compact.

7. *Table A1.* This remark is not very clear to us. While size bias of asymptotic tests typically does grow with N, this is not the case for bootstrap tests (see, *e.g.*, Fachin, 2007, Westerlund and Edgerton, 2007). Note that in our case small N results should be assessed keeping in mind that stability statistics are very sensitive to outliers (especially so the $SupF$ test).

8. *Section 4 (empirical application).* This topic includes many specific points. We will discuss them in turn.

"*the figures don't show much evidence in favor of breaks*". We believe this remark to be mostly due to the format and scale of the plots. From a

different set of plots (included in this reply, see below) the hypothesis of a weaker link between savings and investments in the last years of the sample, when a widespread move towards financial liberalisation took place, appears plausible enough to be tested.

"*you say that the fact that while the individual tests generally don't reject the null of no cointegration (...)*" . Westerlund has clearly been misled by an unsatisactory presentation of our results. In this paper, relying on the results of the companion paper Di Iorio and Fachin (2007), we test the hypothesis that a cointegrating relationship *assumed* to exist is stable (H_0 : "no breaks in the cointegrating relationship").

"*For the panel tests to reject, it is sufficient that there is evidence* [in favour of rejection] *for a single unit*". This is not the case. Since our panel tests are based on the mean and the median of the individual tests their outcome reflects the evidence provided by the mass of their distribution. Hence, the failure of *most* individual tests of a panel to reject when a panel test does can only explained by low power of the former (which is an established fact since Gregory, Nason and Watt, 1996).

dynamics, 1960-2002. S/Y (solid line) and I/Y (dotted line).

Savings (S) and Investments (I) to GDP (Y) ratios dynamics, 1960-2002.
S/Y: solid line; I/Y: dotted line.

Savings (S) and Investments (I) to GDP (Y) ratios dynamics, 1960-2002.
S/Y: solid line; I/Y: dotted line.

2   Minor comments

Most of the comments are very helpful suggestions on how to improve the presentation, which have been followed in the revision of the paper. The only points requiring a response here are the following (numbers as in Westerlund's comment):

6. *Page 4, remark (i) (definition of the breakpoint estimate)* This question is probably partly due to an infortunate typing mistake in the paper. The breakpoint estimate for unit i is actually defined as $\widehat{t}_i^b = \arg\max(\widehat{F}_i)$, *i.e.,* the observation corresponding to $SupF$. However, as discussed

in the paper, rather than using this estimate in the bootstrap algorithm we replace it with the median of the estimates for all units, $\widetilde{t}_i^b = median(\widehat{\mathbf{t}}^b)$, where $\widehat{\mathbf{t}}^b = \left[\widehat{t}_1^b \widehat{t}_2^b \ldots \widehat{t}_N^b\right]$.

7. *Page 6, line 2 beneath (2): "Don't you mean strictly exogenous? u_{it}^x and u_{it}^y it are uncorrelated"* In the DGP at hand, this is indeed the case. Here, however, the issue is on the condition required for single-equation methods to be valid, which is weak exogeneity.

8. *Page 8, case 2: "If you want to see the effect of increasing T you should increase it while keeping N fixed".* Here the point is that for computational reasons we could handle only very small cross-section sample sizes. Hence we considered two cross-section sample sizes, 3 and 5; the latter is present in the simulations with $T = 50$ as well, thus allowing comparisons exactly as that suggested by Westerlund. The other ($N = 3$) is included to allow a comparison of the results obtained with a large time sample and with different cross-section sample sizes.

9. *Page 10, section 3.2, line 1: "What is meant by the "speed of adjustment of the DGP's"?"* Speed of adjustment is determined by the autoregressive coefficient of the noise of the Y equation, ϕ. The closer this is to 1, the slower the adjustment (in the limit, $\phi = 1$ implies no cointegration).

12. *Page 21, lines 15 to 18: "The p-values* [of the asymptotic tests] *you get are much more likely to be due to size distortions than to low power".* Our conjecture is based on the fact that asymptotic tests are well known in the literature to suffer from low power: even with $T \geq 100$ Gregory, Nason, Watt (1996) conclude that "Power is poor unless the stable root is small".

3 References

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/

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

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

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

Westerlund, J. and D. L. Edgerton (2007) "A panel bootstrap cointegration test" *Economics Letters,* 97: 185-190.