Response to Referee report 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 the Referee for his accurate report including many helpful suggestions. We will now try to answer in turn to the various questions raised. Obviously, all the following discussion has been included in the revised version of the paper.

1 Monte Carlo design

The choice of the parameters was made on the basis of two distinct rationales.

First of all, *speed of adjustment*. This affects in the same way the asymptotic and the panel tests: both will deliver better results when the long-run relationship stands out more clearly from the data, *i.e.* when the speed of adjustment is higher (ϕ closer to 0). Since the primary aim of our experiment is assessing the scope for improving over pure time series testing using a panel approach we can then fix the $\phi's$ with no loss of generality. A natural choice are the values adopted by Pesaran (2007) in a recent important work on the closely related issue of panel unit root tests.

Second, the parameters governing the panel structure. These are the *factor loadings* and the *noise variances*, which have been indeed chosen on the basis of real world data, although we did not explain this in the paper. More specifically:

Factor loadings ($\gamma_i, i =$unit index). In the panel of european economies used in our empirical application the maximum correlation across units of the log differences of savings (the right-hand side variable) and of the residuals of the Feldstein-Horioka equations are respectively about 0.62 and 0.65. We then set $\gamma_i \sim Uniform(-1, 6)$, which yields a cross-correlation close to 0.65, for both variables.

Noise variances (σ_i^2, $i =$unit index). In the same dataset we find that after normalising on the cross-section average in ten out of 12 cases 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 $[0.38, 1.49]$. We thus adopted Pesaran's (2007) choice, $\sigma_i^2 \sim Uniform(0.5, 1.5)$, which seems to be reasonably representative of empirical data sets

1

2 Comparing the power of tests

Since we use the bootstrap distribution directly to compute the p-values of the empirical statistic under the null hypothesis the size-adjustment would imply 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.

Clearly, using the empirical size, rather than the nominal level α, will lower the power of oversized tests and increase that of undersized ones, making different tests directly comparable. However, as stressed by Horowitz and Savin (2000), who go as far as describing size-adjusted power as "irrelevant for empirical research", this advantage comes at the expense of the relevance of the Monte Carlo simulations for actual empirical work. We shall now argue that in our case the benefits of size-adjustment would not be worth this cost. The comparisons made in the 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 it happens, 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 increse 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 power.

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.

For instance, in the empirical application of this paper we write:

"Here the evidence for rejection is weaker, with p-values between 10% and 15% for the L_c and $MeanF$. However, recalling (...) 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".

And similarly, in Di Iorio and Fachin (2007):

"We thus turn to the bootstrap panel cointegration tests (...) to correctly evaluate [these p-values] we should keep in mind that (...) the possible breaks are widely dispersed across units. From our simulations (...) we know that in these circumstances our panel cointegration tests may be severely undersized: hence, [they] should be regarded as significant".

Clearly, simulation results on size-adjusted power would not been able to provide a similar help in interpreting these empirical results.

To conclude, we can easily compute size-adjusted power, but we believe in this case a careful reading of raw power to be more useful.

3   Comments by page and line

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:

p. 4 - l. 4 by *mean estimation error* we refer to the distortion. We made this point clearer in the revised version.

Fig. 1 The peak is on -1. We remade the figure so to make this clear.

p. 5 - last 2 lines and the following This remark is fully correct. Indeed, our discussion of the Monte Carlo Data Generating Process (DGP) was not very clear. We started from a general form in order to highlight the links with the DGP's used in important papers, such as Gonzalo (1994), but failed to state clearly how we restricted it by the assumption of exogeneity. X is indeed not breaking.

eq. (4). This was indeed an infortunate typo.

4 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/

Gonzalo, J. (1994) "Five alternative methods of estimating long-run equilibrium relationships" *Journal of Econometrics* 60: 203 -233. 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, pp. 375-389

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