

# Testing for unit roots on heterogeneous panels : A sequential approach

Pierre-Yves HENIN<sup>¶</sup>, Philippe JOLIVALDT<sup>‡</sup>, Anh NGUYEN<sup>‡</sup>

Revised July, 2001

## Abstract

There is a growing trend of criticism against the use of panel data unit root test for assessing hypotheses such as the purchasing power parity. The usual argument of a gain in power with respect to univariate unit root tests is not relevant as different nulls are involved when testing on panel data.. In the context of a comparative, multi-country, study, inference based on individual unit root tests suffers mainly from a huge size distortion, even more than from low power.

When the null hypothesis is -as usually in comparative research- a number of countries for which the variable of interest follows a unit root process, we propose to adopt a sequential strategy, as a way to combine a rigorous control over the size with the search for satisfactory power. We show how usual statistics have to be adapted, and we illustrate the implementation of this strategy both through simulation and through an empirical application to the PPP hypothesis within the OECD countries.

**Keywords :** panel unit root tests, multiple testing, sequential tests, PPP.

JEL classification : C12, C23, F31

---

<sup>¶</sup>EUREQua - Université de Paris-I, CEPREMAP

<sup>‡</sup>EUREQua - Université de Paris-I

<sup>‡</sup>EUREQua - Université de Paris-I

# 1 Introduction

Many studies use panel unit root test on macroeconomic, multi-country data. The purchasing power parity (PPP) is the hypothesis of interest in many cases (e.g. Coakley and Fuertes, 1997, Frankel and Rose, 1996, Heimonen, 1999, Higgins, 2000, Mac Donald, 1996, Oh, 1996, Wu, 1996) but the panel unit root tests are also used to assess the stationarity of the inflation rate (Culver and Papell, 1997), of interest rates (Wu and Zhang, 1996), of the output per head (Fleissig and Strauss, 1999) or the current account (Wu, 2000).

The use of panel tests for unit roots is usually motivated by the fact that they provide a gain in power. Maddala (1999) convincingly criticizes this argument, owing to the irrelevance of comparing power between tests which do not have the same null hypothesis. Furthermore, these tests differ according to the degree of heterogeneity allowed between the individuals, either under the null or the alternative.

Typically, the Im, Pesaran and Shin (1996) test, thereafter IPS, and the Maddala and Wu (1999)  $P_{\text{rest}}$  rest on the alternative that at least one of the individual variable follows a stationary process thus implying that unit root behaviour is not of common characterization of the individual series. This makes the issue very dependant on the sample composition. Thus, researchers frequently argue that the null is rejected (or not) when some country is discarded (e.g. Japan, in Choi, 2001, study of PPP). This practice indicates that the IPS's null is probably not the main hypothesis of interest for many researchers. Of course, it is legitimate to retest for PPP excluding one country, e.g. Japan from the panel. The error comes from not adjusting the critical values for changes in actual type-I error.

In this paper, we propose to consider explicitly that researchers are mainly interested in assessing comparative persistence between groups of individuals -or groups of countries. The typical question is : for how many countries does the inflation (or the unemployment) rate follow a unit root process ? Or, subsequently, can we classify the members of the panels in two groups, for which a common unit root is respectively rejected and non rejected ? Such a design of the test paves the way for further analysis, either positive : e.g. which institutional difference accounts for those different dynamics ? -either normative-. e.g. In which country are policy changes required to correct for persistent inflation or unemployment divergences ? In order to have a handle on this question, we propose to follow a sequential strategy,

progressively eliminating from the panel those countries with the "less persistent" variables. This recursive strategy requires the use of an adapted, iterated version of panel unit root statistics, which have to account for both combination and truncation effects.

The remainder of the paper is organized as follows. Section 2, shows that the main pitfall in using individual unit root test is not lack of power, but size distortion due to the neglect of the "data mining" involved by repeating a test for different countries. Section 3 develops our new sequential, inference strategy. Section 4 indicates how panel unit root tests have to be extended to account for the local nulls encountered under this strategy. Section 5 provides a first assessment of the approach, considering actual size and power of the statistics. Section 6 completes this evaluation by simulating the implementation of our inference strategy in order to determine the number of unit-root-countries within a panel. Section 7 introduces a bootstrapped version of the statistics in order to cope with the lack of independence between individuals. Section 8 provides results from application to testing PPP on a sample of OECD countries. Section 9 concludes the paper.

## 2 Pitfalls with the univariate ADF approach

In many cases, the matter of interest is the relative persistence of a variable among a set of individuals (countries, regions or sectors...) especially, the determination of a subset of countries in which this variable display "full persistency", associated to a unit root dynamics. The spontaneous strategy is then to run for every individual, an ADF (or equivalent) unit root test. A typical conclusion is that we are able to reject the unit root null for a small part of the sample (e.g. 3 or 5 over 20). A usual view is that this result comes from the lack of power of ADF test. Actually for  $T = 100$  and an autoregressive root of 0.9 under the alternative the percentage of rejection (at the 5% level) is limited to 33%, reaching 88% for a root of 0.8 under the alternative. This is a common argument for switching to an alternative strategy, considering the sample as a panel.

However, researchers are rarely aware of a more severe pitfall of the spontaneous strategy, which results from a huge size distortion. Actual type I error, as the risk of rejecting the (true) null for every individual, is clearly much lower than the nominal size of spurious rejection for any individual considered isolately. Using standard nominal levels to derive statements such

as “we reject a unit root in the unemployment rate for only two of twenty countries” would lead to spurious inference as it involves data mining, in a way that is very similar to the one discussed by Lovell (1983).

Let us consider the following illustration of this problem. We simulate 1000 replications of  $N = 20$  random walks of length  $T = 100$ : We reject the unit root for one country if the smallest of the  $t_{ADF}$  statistics  $t_{ADF}^{(1)}$  exceeds the usual  $\alpha$ -level critical value. We reject it for two countries if the second smallest exceeds the same value, ...etc... The empirical percentages of rejection, i.e. the actual size of the test, are reported on table one.

rank n	20	19	18	17	16	15	n < 15
size $\alpha$							
0.05	0.64	0.17	0.02	0.01	0.00	0.00	0.00
0.10	0.88	0.60	0.32	0.13	0.04	0.01	0.00

Table 1 : Empirical percentage of rejection, univariate ADF

At the 5% level, the actual size is 64% for the “least persisting country” with the smallest empirical  $t_{ADF}$  statistic, and 17% for the “second least persisting”. Then, the size collapses much below the nominal level, practically disappearing from the fourth country. At the 10% level, we observe a similar profile. Thus using constant critical values when testing among a set of country strongly departs from a major requirement of classical testing : to control a constant type I error.<sup>1</sup> The naive strategy is clearly biased toward finding a “low percentage of rejection”, against both of the alternative of no rejection (the consequence of size distortion if the data are  $I(1)$ ) and of a high percentage of rejection (the consequence of low power if data are  $I(0)$ ). Thus, we may wonder if the high frequency of studies relating “a few rejections of the unit root null” is really a feature of international data, or rather a mechanical by product of the testing strategy.

---

<sup>1</sup>Chow and Denning (1993) make a similar point, examining the type I error from comparing the maximum of a series of (modified) variance ratio statistics to a standard normal critical value.

### 3 An alternative approach : test combination and a sequential strategy

A second objection against the argument in favor of panel unit root tests as providing a gain of power is given by Maddala and Wu (1999) : it makes no sense to compare tests with different null hypotheses. Let us thus compare the three sets of hypotheses associated with three usual panel unit root tests.

Case 1 (Levin and Lin)

$$\text{Assuming } \frac{1}{2}_1 = \frac{1}{2}; \delta_i; \quad H_0 : \frac{1}{2} = 1; \\ H_a : \frac{1}{2} < 1$$

Case 2 (Im, Pesaran and Shin)

$$H_0 : \frac{1}{2}_i = 1; \delta_i \\ H_a : \rho_i; \frac{1}{2}_i < 1$$

Case 3 (JLR test, Taylor and Sarno, 1998)

$$H_0 : \rho_i; \frac{1}{2}_i = 1 \\ H_a : \delta_i; \frac{1}{2}_i < 1$$

Despite its variety, we are not sure that this set of cases matches the demand of researchers. Commonly, the hypotheses of N or 0 occurrence of a non stationary series among N countries are only limiting ones. As revealed by the (more or less formal) comments in many papers, a deeper interest lies with the number and the identity of those countries for whom the null is not rejected. The previous exercise decisively shows that the naive approach of individual testing with constant critical values is flawed. We therefore propose to adopt an explicit sequential approach.

Let us consider a variable  $y_{it}$  observed for a set of N individuals ( $i = 1:::N$ ) over a length of time T ( $t = 1:::T$ ); and  $t_i^{ADF}$  the individual ADF t-statistics for these individuals. After arranging these individuals according to the value of their  $t_i^{ADF}$ , we propose to search for tests supporting the partition of the sample between GS (n) and GU (n),

$$f_{GS(n)}; i = 1; :::; N; j = n \\ f_{GU(n)}; i = N; j = n + 1; :::; N$$

such that the "local null" hypothesis<sup>2</sup> :

$$H_0(n) : \frac{1}{2} \rho_i = 1; \text{ } 8i \in 2 \text{ GU}(n)$$

is not rejected against  $H_a : \rho_i : \frac{1}{2} \rho_i < 1$   $H_a(n) : \rho_i \in 2 \text{ GU}(n) ; \frac{1}{2} \rho_i < 1$   
 but  $H_0(n+1) : \frac{1}{2} \rho_i = 1; \text{ } 8i \in 2 \text{ GU}(n+1)$  is rejected.

$\text{GU}(n)$  ; the set of individuals with a non stationary  $y_{it}$ ; is then defined as the largest one for which the "local null"  $H_0(n)$  is not rejected.

This strategy is initialized for  $n = N$ : The local  $H_0(N)$  is thus the usual null of panel unit root test. In case of no rejection, there is no reason to proceed further as we may conclude that the set of countries with a stationary  $y_{it}$ ;  $\text{GS}(n)$ ; is void. However, if  $H_0(N)$  is rejected, we propose to consider formally the following local null  $H_0(N - j - 1)$ : In case of rejection, we conclude to one "stationary country" and  $N - j - 1$  non stationary. In case of non rejection, the process is continued, as far as required.

Thus the ultimate outcome of our strategy is a partition of the individuals, beforehand arranged according to their "persistence level", between those with a stationary  $y_{it}$ ; and those with an integrated  $y_{it}$ :

Having thus clearly defined our inference strategy, we may develop the design of proper tests.

## 4 Panel unit root tests and their sequential extension

If we consider only the heterogenous panel case (allowing for different  $\frac{1}{2} \rho_i$ ), with the null and alternative hypothesis as defined in case 2, panel unit root tests involve a dimension of test combination. So, the IPS  $\bar{t}$  statistic is defined as a properly centered and standardized mean of the individual  $t_i^{\text{ADF}}$ : This "test combination" dimension is even clearer for the Fisher-Pearson  $P_{\tau}$  statistic advocated by Maddala and Wu (1999)<sup>3</sup>

<sup>2</sup>Formally, a "local null"  $H_0(n)$  is parametrized by  $n = N_U$ ; the size of the subset of individuals with a unit root behavior.

<sup>3</sup>As this paper was being analyzed, we became aware of Choi's (2001) contribution which provides asymptotic results for the distribution of different combination statistics.

$$P_{\alpha} = \prod_{i=1}^N 2^{-\log p_i} \quad (1)$$

where  $p_i$  stand for the  $p_i$  value of the individual ADF (or any other appropriate individual statistic). Using this "combination argument" we may consider as a relevant (if not efficient !) panel unit root test the minimum DF statistic,  $t^m$  defined as :

$$t^m = \text{Min}_{i \in \{1, \dots, N\}} t_i \quad (2)$$

The distribution of the  $t^m$  statistic follows from the observation that :

$$\Pr(t^m < t_{\alpha}) = 1 - \Pr(t_i \geq t_{\alpha}) \quad (3)$$

thus, assuming the  $t_i$  distribution to be identical<sup>4</sup> and independent<sup>5</sup>

$$\Pr(t^m < t_{\alpha}) = 1 - (1 - \alpha)^N \quad (4)$$

Hence, the critical values at the  $\alpha$  level to which one should compare the empirical  $t^m$ <sup>6</sup>

$$\alpha(\alpha; N) = 1 - (1 - \alpha)^{1=N} \quad (5)$$

For instance, for  $\alpha = 10\%$  and  $N = 10$ ; the critical value to which the  $t^m$  statistic refers is not  $t_{0.1}$  (-2.58) but, as  $\alpha(0.1; 10) \approx 0.01$ ;  $t_{0.01}$  (-3.51).

This "test combination" mechanism defines a first dimension of the required adaptation of statistics in order to deal with our sequential strategy. However, our strategy also implies to consider truncated distributions, which introduce a second dimension of adaptation of the statistic.

<sup>4</sup>For  $t_i^{\text{ADF}}$ ; at finite distance, identical distribution would require a common  $k_i$  lag order.

<sup>5</sup>Independence is a common assumption for the definition of panel unit root test (e.g. IPS, 1997). We will consider later the consequence of the lack of independence.

<sup>6</sup>This result, as the choice of the "minimum statistic" for test combination traces back to Tippet, in 1931. See ESS (1988, p.216) and Folks (1984).

## 4.1 A marginal Dickey-Fuller statistic

Let us consider first the case of an iterated version of the minimum  $t^m$ ;  $t^m(n)$ ; which can be labelled the marginal Dickey-Fuller for a subset of  $n$  individuals :

$$t^m(n) = \text{Min}_{i > N_i n} t_i$$

**Proposition 1** : The  $t^m(n)$  marginal DF statistic under the local null  $H_0(n)$  is distributed over  $]t_n; +1]$  with a CDF derived from the Dickey-Fuller CDF by the following transform

$$\Pr \{ t^m(n) < t_n^{\text{DF}} \} = 1 - (1 - p(t_n))^n \quad (6)$$

with  $p(t_n) = \Pr \{ t^{\text{DF}} < t_n \}$ ; and  $t_n$  is the  $t^{\text{ADF}}_i$  statistics of the  $(N_i - n_i - 1)^{\text{st}}$  individual, i.e. the one eliminated at the previous step of the sequential procedure.

**Proof.** : The  $t^m(n)$  test, on the minimum of the  $t_i$ ;  $i > N_i n$  is equivalent to a joint test on these  $n$   $t_i$ ; assumed independent. According to the Tippett formula (equation 4),  $\Pr \{ t^m(n) < t_n^{\text{DF}} \} = 1 - (1 - \mathcal{E}(n))^n$  where  $\mathcal{E}(n)$  stands for the probability of a  $t$  value lower than  $t_n^{\text{DF}}$  in the distribution truncated at  $t_n$ : It is straight forward to express  $\mathcal{E}(n)$  as : ■

$$\mathcal{E}(n) = \frac{1 - p(t_n)}{1 - p(t_n)}$$

which substituted in the previous expression, gives (6). ¥

Equation (6) makes clear that the distribution of the marginal  $t^m(n)$  depends on both  $n$  and the truncature parameter  $t_n$ ; but not directly on the full sample size,  $N$ : This equation is also informative by clearly disentangling the two correction factors, respectively for the combination effect and for the truncation effect.



## 4.2 An iterated Fisher-Pearson $P_{\tau}$

Another panel statistic of interest is the Fisher-Pearson  $P_{\tau}$ .<sup>7</sup> The adaptation of the standard  $P_{\tau}$  (equation 1) to our sequential strategy is straightforward. Let us introduce :

$$P_{\tau}(n; t_n) = \frac{1}{2} \sum_{i=N_j+1}^n \log p_i(t_n) \quad (7)$$

where  $p_i(\cdot)$  stands for the  $p_i$  value evaluated according to the truncated distribution, i.e.

$$p_i(t_n) = \frac{p_i \cdot p(t_n)}{1 - p(t_n)} \quad (8)$$

We may conjecture that, with the  $p_i$  value corrected for the truncation effect,  $P_{\tau}(n; t_n)$  converge towards a  $\hat{A}_{2n}^2$  distribution independant of  $t_n$ : Simulation evidence will support this conjecture.

## 4.3 An iterated IPS $P_{\bar{\tau}}$ statistic

We may now consider the required extension of the IPS statistic. Let us define the iterated IPS  $P_{\bar{\tau}}(n; t_n)$  as :

$$P_{\bar{\tau}}(n; t_n) = \frac{\frac{1}{n} \sum_{i=1}^n \bar{t}_n \cdot E \left[ t_{i > N_j+1} ; \frac{1}{2} \right]}{\text{Var} \left[ t_{i > N_j+1} ; \frac{1}{2} \right]} \quad (9)$$

with  $\bar{t}_n = \frac{1}{n} \sum_{i=N_j+1}^n t_i$

and  $E(t_{i:::})$  and  $\text{Var}(t_{i:::})$  denoting respectively the esperance and variance<sup>8</sup> of the individual  $t_i$  statistics for  $i^2\text{GU}(n)$ :<sup>9</sup>

<sup>7</sup>The notation  $P_{\tau}$  is due to Pearson, and initially applied to bilateral statistics -see Folks (1984)- Becker (1997) compares  $P_{\tau}$  with other possibilities of  $p_i$  value combination. Psaradakis (2000) focuses on p-value adjustments required when combining multiple tests for non linearity.

<sup>8</sup>Setting  $n = N$ , equation (10) involves only the unconditional (w:r:t:i) expectation and variance and therefore coincides with the usual definition of the IPS statistics.

<sup>9</sup>Alternatively, we may define iterated versions of other panel unit root statistics, as the JLR  $\tau_i$  min one. However, the use of these statistics, involving different null hypotheses, would imply an adaptation of our sequential strategy. Similarly, our approach may be modified as to take a stationarity hypothesis as the "local null".

We will proceed further considering these three tests statistics : the marginal DF,  $t^m(n)$  ; the iterated  $\hat{p}_s(n; t_n)$  and the iterated IPS  $\hat{p}_T$ : Inference based on the naive DF will also be reported as a benchmark.

The following table sums up the features of these four test statistics.

Test	Statistic	Critical value	Correction for	
			Composition effect	Truncation effect
marginal DF	$t^m(n)$	$t_{\hat{\theta}}^{DF}$	adjusting the level $\hat{\theta}$	adjusting the level $\hat{\theta}$
iterated Fisher-Pearson	$\hat{p}_s(n; t_n)$	$\hat{A}_{\hat{\theta}}^2(2n)$	normalizing the $p_i$	centering the $p_i$
iterated IPS	$\hat{p}_T(n; t_n)$	$t_{\hat{\theta}}^N$	averaging the $t_i$	centering the $t_i$
naive DF	$t_i^{ADF}$	$t_{\hat{\theta}}^{DF}$	no correction	no correction

Table 2 : Comparison of the test statistics

## 5 Distribution under the null and computation of the critical values

The empirical implementation of our strategy requires to compute critical values by simulating the distribution of the statistics under the null hypothesis. As previously noticed, we have only conjectured that the  $\hat{p}_s(n; t_n)$  and  $\hat{p}_T(n; t_n)$  would asymptotically behave as  $\hat{A}^2(2n)$  and  $N(0; 1)$  respectively; and it is important to assess the validity of this conjecture and its empirical relevance at finite distance. It would also be interesting to check the empirical behaviour of the marginal DF with respect to its theoretical distribution characterized by proposition 1.

A difficulty in designing the simulation exercise is to ensure that parameters  $n$  (number of individual series) and  $p(t_n)$  (the truncation) parameter are independent under the local null  $H_0(n)$ :<sup>10</sup> The simulation protocol is thus defined to satisfy this independence requirement.

<sup>10</sup>This independence condition also ensures that the critical values provided here are independent of the sample size  $N$ : A related remark is that the correlation between  $n$  and  $p(t_n)$  in an empirical distribution increases with the weight of the alternative.

The critical values reported on table 3, are is derived from the application of our test over 30.000 replications of a randown walk of length  $T=100$ .

In every case, the full size of the panel is fixed at  $N=20$ . Results are reported here with only a purpose of illustration, as for finer grid of  $p(t_n)$  tabulations are required for an actual implementation of our methodology. Critical values (at 5%) are reported on table 3. One can notice, by a line, the impact of composition effect and, by column, the incidence of the truncation effect. For large  $n$  and a small truncation parameter, we get, more conservative, smaller critical values (than the  $t_1^{DF} = -2.89$ ). However, for most of the combination  $(n; p(t_n))$ ; we get larger critical values allowing easier rejections. This table illustrates the extent of the required correction to be used by a researcher inferring on the relative persistence of a variable on a panel of countries.

n	20	18	15	10	5	2	1
0.00	-3.90	-3.86	-3.84	-3.68	-3.50	-3.14	-2.87
0.015	-3.29	-3.28	-3.29	-3.24	-3.20	-2.98	-2.77
0.05	-2.86	-2.86	-2.85	-2.84	-2.83	-2.72	-2.60
0.10	-2.57	-2.57	-2.56	-2.56	-2.54	-2.49	-2.41
0.25	-2.09	-2.09	-2.08	-2.08	-2.07	-2.04	-1.99
0.50	-1.55	-1.55	-1.46	-1.55	-1.52	-1.53	-1.50
0.75	-0.99	-0.99	-0.99	-0.98	-0.97	-0.97	-0.95
0.90	-0.406	-0.406	-0.402	-0.405	-4.29	-0.394	-0.381
0.95	-0.067	-0.067	-0.037	-0.065	-0.060	-0.055	-0.042
$t_1^{DF}$	-2.89	-2.89	-2.89	-2.89	-2.89	-2.89	-2.89

Table 3 : Critical values of the marginal DF  $t^m$  statistic.

Similarly, the table 4 reports the critical values of the iterated Fisher-Pearson  $\hat{F}_n^*$ : Notice that the first value  $\hat{F}_n^*(20; 00)$  is the one applicable to the full panel (as suggested by Maddala and Wu (1999)). This computed value is surprisingly close to the standard  $\hat{A}_n^2(40)$ : We notice also that the simulated values for the iterated  $\hat{F}_n^*$  stay very close to the theoretical  $\hat{A}_n^2(2n)$ : These results support the practice of using the standard tabulated values directly for inference.<sup>11</sup>

<sup>11</sup>The practical relevance of this success is however limited, as this result is obtained assuming independant probabilities, a feature not likely to be encountered on actual panels.

n	20	18	15	10	5	2	1
p(t <sub>n</sub> )							
0.00	55.54	50.95	44.09	31.03	18.4	9.49	6.07
0.015	55.72	50.73	43.94	31.51	18.23	9.54	6.08
0.05	55.8	50.72	43.77	31.18	18.17	9.7	5.89
0.10	55.75	50.82	43.67	31.32	18.42	9.7	6.06
0.25	55.68	51.04	43.39	31.2	10.06	9.38	6.04
0.50	55.81	50.88	43.64	31.24	18.14	9.43	5.9
0.75	55.52	50.74	43.72	31.36	18.11	9.42	5.88
0.90	55.69	50.72	43.11	31.33	18.11	9.44	6.11
0.95	54.93	50.14	42.8	30.94	17.86	9.43	5.99
$\bar{A}^2(2n)$	55.8	50.99	43.8	31.41	18.31	9.49	5.99

Table 4 : 95% critical value, iterated  $\hat{\rho}$

Table 5 similarly reports (5%) critical values of the iterated IPS statistic. Again, the first value, obtained for  $n = 20$  and  $p(t_n) = 0$  is the critical value applicable for the usual IPS test on the full panel, and it is found to be very close to the normal one. The iterated value is very weakly sensitive to  $n$ ; at least for  $n > 2$ ; and thus to the pure composition effect. It is more sensitive to the truncation effect, decreasing when  $p(t_n)$  increases, especially for small  $n$ : Thus, the conjecture that  $\hat{\rho}_T(n; \cdot)$  converges to a  $N(0, 1)$  when  $n$  increases is verified, the convergence being slower the larger is  $i$ ; thus the truncation.

n	20	18	15	10	5	2	1
p(t <sub>n</sub> )							
0.00	-1.65	-1.62	-1.63	-1.64	-1.63	-1.62	-1.58
0.015	-1.61	-1.62	-1.64	-1.61	-1.58	-1.57	-1.57
0.05	-1.61	-1.60	-1.57	-1.60	-1.59	-1.53	-1.50
0.10	-1.59	-1.60	-1.59	-1.59	-1.55	-1.49	-1.43
0.25	-1.59	-1.57	-1.56	-1.56	-1.50	-1.41	-1.26
0.50	-1.54	-1.55	-1.56	-1.52	-1.48	-1.36	-1.15
0.75	-1.55	-1.55	-1.54	-1.52	-1.46	-1.31	-1.10
0.90	-1.55	-1.53	-1.53	-1.49	-1.44	-1.25	-1.04
0.95	-1.54	-1.54	-1.49	-1.51	-1.44	-1.26	-1.03
$t_{0.05}$	-1.64	-1.64	-1.64	-1.64	-1.64	-1.64	-1.64

Table 5 : 5 % critical values, iterated IPS  $\hat{\rho}$

According to these numerical results, the distributions of the proposed iterated test statistics appear well behaved, which supports their use as practical tools in empirical studies.

## 6 An empirical evaluation of the sequential approach

The empirical evaluation of the proposed sequential testing procedure has to be performed in two steps. In both cases, pseudo samples are constructed through simulation of a variety of data generating processes characterized by the local null. Actually, we consider, for  $N = 20$ ; six values of  $N_u$ ; the number of individual series with a unit root,  $N_u = 0; 1; 5; 10; 15; 20$ : Using the full panel ( $n = 20$  for  $N = 20$ ); the rejection frequencies give the empirical size, (for  $N_u = 20$  in the DGP's) and measures of power (for  $N_u < 20$ ) of panel unit root statistics. When we proceed with the sequential elimination of the "most stationary" individuals ( $n = 19; 18; \dots; 1$ ); the empirical size is obtained for the local null (i.e.  $N_u = n$ ) and power against this local null for DGP's with  $N_u < n$ :

### 6.1 Assessing size and power of the statistics

Figure 1 reports results for the 5% level and an autoregressive root under the alternative  $\frac{1}{2} = 0.8$ ; which ensures a fairly good power (88% for  $T = 100$ ) for the ADF on individual series.

To read these graphs, we notice that for  $n = N_u$ , the rejection frequencies measure empirical size, and are expected to be constant and close to the nominal size (5% in the current exercise). For  $n > N_u$ ; and therefore everywhere for the curve  $N_u = 0$ ; these frequencies denote power, for which values close to one are desirable.

Insert here Figure 1 : Empirical percentage of rejection under alternative hypothesis :  $\frac{1}{2} = 0.8$ ;  $\alpha = 5\%$ ;  $N = 20$ ;  $T = 100$

A first result from this simulation exercise is that the three consistent statistics have good size properties. A non negligible distortion arises only

when the number of unit root individuals equals the number of individuals retained in the panel ( $n = N_u$  or  $n \approx N_u$ ). The power of the marginal Dickey-Fuller,  $\hat{\rho}^{\min}$ , is low for this parametrization of the alternative, while the iterated IPS  $\hat{\rho}$  and especially the iterated Fisher-Pearson  $\hat{\rho}_F$  perform much better. For instance, assume there are only 5 unit roots in the DGP ( $N_u = 5$ ): We need a panel of 12 individuals (hence 7 following the alternative) for the  $\hat{\rho}^{\min}$  to give a power of 80%. The same power is reached with  $\hat{\rho}$  for panels of 8 (3 following the alternative, in probability) and with the  $\hat{\rho}_F$ ; for panels of 6 or 7.

Only at the border case, for  $N_u = 19$  and  $n = 20$ ; is the power of the marginal DF greater than the power of the iterated  $\hat{\rho}$  and  $\hat{\rho}_F$ . An important result is the excess power of the iterated Fisher-Pearson with respect to the iterated IPS. This statistics therefore cumulates advantages with respect to  $\hat{\rho}$ : it is easier to compute, it converges rapidly towards a standard  $\hat{A}^2$ ; and it has the best power. At the opposite, although this exercise confirms the correct size properties of the marginal DF, its lack of power is a decisive drawback.<sup>12</sup>

Figure 1 also illustrates the pitfall with the naive ADF as its size lies close to zero for much of the range  $n < N_u$ : However, it presents the largest power for large  $n$  and  $N_u$ :

It is however disputable to argue from this power to advocate the use of naive ADF in an inference strategy, as it is mainly spurious, as a by product of size distortion. For instance, with 19 unit roots in the DGP, tests using the naive ADF on a (truncated) sample of the 18 "less stationary" individual would reject the (true) null of 18 unit root individuals in more than 60% of the cases.

A replication of this exercise, with an autoregressive root  $\frac{1}{2} = 0.9$  under the alternative has been performed, but not reported here. It decisively argues against the marginal DF, the power of which never reaches 50%. The reduction of power of the iterated IPS and especially the iterated Fisher-Pearson is serious, and the performances of these statistics are comparable for small  $N_u$ :

The paradoxical behavior of the naive Dickey-Fuller is confirmed, as it is the only statistic to keep some power in the neighborhood of the global null (i.e. for  $N_u \approx 20$ ):

---

<sup>12</sup>Choi (2001, p. 253) also indicates having discarded a Tippett statistic owing to its poor finite sample performance.

## 6.2 Simulating the sequential strategy

In the previous exercise, we reported the results from considering all the possible values of  $n$ : In the case at hand, our proposed strategy involves an estimate of  $n$  associated to the stopping rule : stop the process of eliminating the "more stationary" individuals as soon as the local null  $H_0(n)$  is not rejected. The most specific evaluation of our inference strategy would thus result from a simulation of this stopping rule strategy.

For every one of the pseudo samples from replicating a DGP, we get an estimate of  $n$  from each of the statistics. We report on figures 2 and 3, panels a to f the distribution of these estimates for different values of  $N_u$ : The other parameters of the experience are fixed,  $N = 20$ ;  $T = 100$ ; the level  $\alpha = 5\%$  and the alternative  $\frac{1}{2} = 0.8$ : As two conciliate lisibility and place saving, the distribution for two statistics are reported on the same figure.

Insert here :

Figure 2 : Distribution of estimated  $n$ , under  $\frac{1}{2} = 0.8$ ;  $N = 20$ ;  $T = 100$ ; IPS and DF statistics

Figure 3 : Distribution of estimated  $n$ , under  $\frac{1}{2} = 0.8$ ;  $N = 20$ ;  $T = 100$ ;  $\hat{p}_u$  and marginal DF,  $t^m$ ; statistics

For the global null, results are quite good for the three consistent statistics, only the naive DF leading to a non negligible frequency of cases of under estimation of  $n$ : The results, however, deteriorate rapidly as the lack of power generates an over estimation of  $n$  by the  $\hat{p}_u$ ; the  $\hat{p}_u$  and especially the marginal DF. Acceptable performances are also obtained in the neighborhood of the global alternative  $N_u \gtrsim 0$ . The estimated values of  $n$  are rather dispersed for intermediate situations, but we may notice that the iterated Fisher-Pearson out-performs the iterated IPS. The distribution of the estimate of  $n$  by the marginal DF become flat for average or low  $N_u$ ; which means that this test statistic is uninformative. The paradox comes from the relative performance of the naive DF statistics, which provides the less disappointing results for intermediate cases, with  $N_u \leq 10$ :

While repeating the exercise for the more unfavorable case<sup>13</sup> (for inference) where  $\frac{1}{2} = 0.9$  under the alternative, we notice a further deterioration of the results for all the statistics. The distribution of estimates of  $n$  using  $\hat{p}_u$  and  $\hat{p}_u$  becomes flatter and more skewed to the left (towards over-estimation).

<sup>13</sup>Tables not reported here, but available upon request.

Things are worse for the marginal DF, but not really better for the naive DF. Although the  $n$  estimates from this statistics are more concentrated, they are also more strongly biased. This result is very important as a warning against the pragmatic argument that the ADF performs not worse than the more sophisticated, consistent estimates of  $n$ ; despite its theoretical drawbacks. The impression may emerge in good power situation (with  $\frac{1}{2} = 0.8$ ): The artefactual nature of the power of naive ADF becomes evident as it bias increases with power.

The very lesson from these exercises may be one of modesty : do not expect too much of test combinations in low power situation, but resist the temptation from a naive strategy offering spurious power at the cost of uncontrollable biases and distortions.

## 7 Accounting for cross individual correlation

The whole set of results presented beforehand was derived assuming independence of individual DGPs. This case is unlikely to appear in any empirical panel, and especially in the case of bilateral real exchange rates such as the one considered when testing for PPP. In the case of correlated individual DGPs, one may consider the use of structural models, coping parametrically with the cross correlations. However, the possibility to cope structurally with cross individual correlations, through a SURE estimator à la Abuaf and Jorion (1990) or a JLR approach (Taylor and Sarno, 1998) is limited to low dimensional panels. The use of Wald test looks a priori promising but, as discussed by Chang (2000), they are bilateral statistics, and thus not suitable for our sequential strategy which implies unilateral truncation.

### 7.1 The bootstrap procedure

We retain the solution advocated by Maddala and Wu (1999), to use the statistics previously defined in the independent case, but with distribution bootstrapped under the null.

In the present case, the bootstrap strategy involves the following steps :

i) estimate an autoregressive model of individual dynamics under the null, as



$$\Phi y_{it} = \sum_{j=1}^{k_i^0} \Phi y_{it-j} + \epsilon_{it}^0 \quad (10)$$

where the lag order  $k_i^0$  is selected according to an information criteria. In order to avoid underparametrization, we use the AIC following Berkowitz and Killian (1996).<sup>14</sup>

ii) Initial conditions for a pseudo sample of  $N$  series are obtained by vector block resampling of the original data. From this drawing, we get the set of conditions  $[y_{i,0}; \Phi y_{i,j-1} \quad \delta_j \quad k_i^0]$  for every individual, respecting the order  $i = 1:::N$ :

iii) Starting with these initial conditions, we built pseudo samples of  $NT$  observations from the DGP (10) and  $T$  vector of errors  $[\epsilon_{1,t}^0:::\epsilon_{i,t}^0:::\epsilon_{N,t}^0]$  drawn with replacement from the vector of empirical residuals  $[\epsilon_{1,t}^0:::\epsilon_{i,t}^0:::\epsilon_{N,t}^0]$ :

iv) Replicating the process (ii, iii)  $R$  times provides a set of  $R$  pseudo samples with the same cross correlations of innovations and initial conditions as the historical one. Applying on this set of pseudo samples the sequential strategy previously described, we get an empirical distribution of the test statistics at every step, conditionnal to  $[n; p(t_n)]$  nuisance parameters of the number of individuals under the local null and the probability of truncature.

Of course these distributions are specific to an historical sample, respecting the initial ordering of the individuals. Thus we cannot give a general tabulation. However, as an illustration we report the  $p$  values obtained for the first case considered in the next section, while testing for PPP on a panel of 16 OECD countries, with monthly data and real exchange rates measured as the bilateral parity vis-à-vis the DM.

## 7.2 An illustration

The consequences of accounting for cross correlation, with the structure of correlation specific to this panel of data, may be assessed by comparing the critical values reported on table 6, panel a for the  $\mathbb{P}_s$  to the ones reported on table 4 for the independence case.<sup>15</sup>

<sup>14</sup>On bootstrapping unit-root statistics, see also Basawa et alii (1991), de Angelis et alii (1997).

<sup>15</sup>Remember that the  $\mathbb{P}_s(n; p(t_n))$  in the independence case does not depend on the sample size,  $N = n$  maximum.

n	16	15	10	5	2	1
p(t <sub>n</sub> )						
0.00	45.84	.	.	.	.	.
0.025	.	43.23	.	.	.	.
0.05	.	45.04	.	.	.	.
0.10	.	44.99	.	.	.	.
0.25	.	44.56	34.32	.	.	.
0.50	.	.	32.24	18.48	.	.
0.75	.	.	.	18.64	10.00	5.73
0.90	.	.	.	18.16	9.89	5.36
0.95	.	.	.	.	9.57	5.57

Panel 6.a : The iterated  $\hat{\rho}_s(n; p(t_n))$

n	16	15	10	5	2	1
p(t <sub>n</sub> )						
0.00	-2.21	.	.	.	.	.
0.025	.	-2.12	.	.	.	.
0.05	.	-1.87	.	.	.	.
0.10	.	-1.96	.	.	.	.
0.25	.	-1.93	-1.79	.	.	.
0.50	.	.	-1.67	-1.52	.	.
0.75	.	.	.	-1.48	-1.19	-0.99
0.90	.	.	.	-1.47	-1.20	-1.00
0.95	.	.	.	.	-1.21	-1.00

Panel 6.b : The iterated  $\hat{\rho}(n; p(t_n))$

Table 6 : Example of critical values  
(The bilateral DM real exchange rate, monthly data)

Although sizeable, the corrections to be introduced are not large. For instance for 15 remaining countries and a probability of truncature of 5%, the critical value of 43.77 in the case of independance becomes 45.04 when accounting for the cross-correlation pattern.

Comparing the effects of accounting for cross-correlation on the iterated IPS statistic, on table 6, panel b (to be compared with table 5), we find greater, and more damaging consequences. For 15 countries, and a truncation at 5%; the 5% critical value is increased from -2.85 to -1.87. For 5 countries,

the simulations do not provide enough cases of truncature to compute a p-value for truncations less probable than 50%. In this situation, the  $\hat{\rho}$  (5; 0:50) of -1.52 in the independence case is left unchanged.

## 8 An empirical application, testing for PPP on a panel of OECD countries

Among the many panel tests for the PPP hypothesis, some studies consider a maximum set of countries while others focus on a small number, representative of a "core" of the international monetary system (the G7 or the G3) or of regional issues (Europe, Latin America, the Middle East ...). We will examine here an intermediate case.

### 8.1 The design of the applications

The applications proposed here are mainly illustrative. We think that the proposed strategy is relevant for a medium-sized, a priori not unduly heterogeneous set of countries, hence the choice of OECD members. Data availability considerations restrict us to considering 17 countries and therefore their 16 bilateral real exchange rates as variables of interest. As various studies have shown (e.g. Coakley and Fuertes, 2000), results are sensitive to both the frequency of data (monthly versus quarterly) and the choice of the country of reference. Real exchange rates are typically found to be less persistent when defined vis-à-vis the DM than when defined vis-à-vis the US \$

In order to illustrate the results of our approach in different situations, we have performed four applications, three of which are reported here.<sup>16</sup> All these applications are relative to the period following the collapse of the fixed exchange rate system (1973.1 - 1998.4). Owing to the presence of several of the EMS countries in the panel, we retain December 1998 as the last observation period in the sample.

Data for the exchange rates are from the IMF, IFS data base, series rf of the average bilateral value. The bilateral rates vis-à-vis the DM are obtained by reporting every national rate vis-à-vis the US \$ to the DM/US \$ rate. Real exchange rates are computed using the CPI series from OECD. We

---

<sup>16</sup>As the two panels on exchange rates defined vis-à-vis the USD implied systematic non rejection, it would have been redundant to present results for both periodicities.

were limited to 17 countries because monthly data over the full period were unavailable for the other OECD countries. For the sake of comparability, we have limited to the same panel the application on quarterly data.

## 8.2 Empirical results

The results of these exercises are presented in tables 7 to 9.

On table 7, we report the results for the bilateral parities vis-à-vis the DM, at a monthly frequency. Applying individual ADF  $t_1$  statistics, the unit root hypothesis is rejected for 9 countries, i.e. by decreasing order of persistence: Italy, the Netherlands, Spain, Denmark, the UK, Finland, Belgium, Switzerland and Portugal.

Insert here table 7 : Empirical results, bilateral DM, monthly

The second line of the table gives the  $p(t_n)$ ; i.e. the truncation probability parameterizing the statistics to be used for testing the iterated local nulls. The  $\hat{P}_n$  presented in the third line have to be compared with their bootstrapped p-values, accounting for cross-correlation as indicated in section 7. We found that the PPP is accepted (i.e. the unit root rejected) for only 7 countries, therefore excluding Switzerland and Portugal.

The iterated IPS statistics are reported on line 5, and the bootstrapped p-values on line 6. On the basis of the iterated IPS statistics, we conclude that PPP is accepted for a group of 14 countries, the only rejecting countries being Austria and Japan.

For the sake of assessing robustness, we may compare these conclusions with the ones based on critical values relevant for the uncorrelated case. According to table 4, the  $\hat{P}_n$  distribution is close to the  $\hat{A}^2(2n)$ : Comparing the empirical  $\hat{P}_n$  to a  $\hat{A}^2$  distribution, we would reject the local null until  $n = 9$ ; with a  $\hat{P}_n = 30:04$ ; exceeding the  $\hat{A}^2(18)$  p-value of 28.9. Similarly, using our general table 5 for iterated IPS under independance, we would reach the same conclusion than with the distribution accounting for cross-correlation.

This first empirical application concludes that the strategy based on the iterated  $\hat{P}_n$  is more conservative than the "naive" one, using the univariate ADF, and especially, than the one based on the iterated IPS.

The second application, reported on table 8, also explores the properties of the bilateral rates vs the DM, but on quarterly data. The ADF loses some power, rejecting the null for only 7 countries (two less than with the monthly data), i.e. in order of increasing persistence : Italy, Spain, the Netherlands, the UK, Denmark, Finland and Switzerland.

Insert here table 8 : Empirical results, bilateral DM, quarterly

Again, the strategy based on iterated  $\hat{\rho}_\mu$  is more conservative, concluding that the PPP is not rejected for only the ...ve "least persisting" countries, Italy to Denmark. We notice, however than using the  $\hat{A}^2(2n)$  approximation, would have allowed for two more cases of rejection ( $n = 11$  and  $n = 10$ ; with  $\hat{\rho}_\mu(10; 0.03) = 33.45$  and a critical value of  $\hat{A}^2(20) = 31.41$ ).

Inference based on iterated IPS also concludes that  $n = 10$ ; therefore leaving 6 countries satisfying the PPP. The cost of accounting for cross-correlation is higher on this case, as, using critical values from the table 5 would have allowed for rejection until  $n = 3$ ; with the same conclusion than the one obtained on monthly data, that PPP holds for 14 countries, excluding only Sweden and Japan.

In our sample, the presence of a unit root in the bilateral rate vis-à-vis the US \$ is never rejected by the ADF. Remembering that  $\hat{\rho}(16,0)$  is the IPS panel unit root statistics, we remark that the null would be rejected assuming no cross-dependency, but not rejected when we account for the cross correlation between any pair of exchange rates. Using similarly the  $\hat{\rho}_\mu(16;0)$  as a full panel unit root statistic, the value of 36.56 did not allow for rejection, when compared either to the bootstrapped p-value (48.83), either to a standard  $\hat{A}^2(32) \approx 46$ .

Insert here table 9 : Empirical results, bilateral USD, quarterly

Every strategy of inference, except the one using IPS and iterated IPS under the independence assumption, concludes to a general non-rejection of unit root, and therefore to the failure of PPP to prevail within the OECD, when the bilateral rates are measured vis-à-vis the US \$.

## 9 Conclusions

This paper has introduced a new strategy, sequentially testing for unit roots in a subset of individuals, -i.e. countries- in a panel. We show that this process involves the use of iterated statistics, which have to be corrected for both composition and truncation effects. Through simulations, we find that an iterated Fisher-Pearson  $F_{\mu}$  performs the best, just followed by our iterated IPS statistic. However, replicating our sequential strategy, including the stopping rule, it is found difficult to estimate the right number of non stationary members of the panel. The cross correlation of variables between countries is found, through bootstrap, to have a sizeable, although modest while limited, impact on the relevant critical values.

Empirical applications to testing for PPP on 17 OECD country members do not deliver results strongly departing from those of univariate ADF tests. The iterated  $F_{\mu}$  test is found to be more conservative than the iterated IPS test, a property not expected a priori.

Further work has to be done to assess the robustness of this result, and confirm whether the gains from the more formal strategy proposed here are generally as modest. Other improvements, unrelated to the core of our arguments, may include starting with more powerful individual unit root statistics, as allowed by the flexibility of the  $P_{\mu}$ ; and exploring the possibility of adapting the bootstrap corrections to account for cointegration (and not only for cross-correlation) between individual series, therefore facing the objections recently raised by Banerjee et al. (2001) and Lyhagen (2000).

## REFERENCES

- Abuaf N. et P. Jorion, 1990, Purchasing power parity in the long run, *Journal of Finance*, 45, pp. 157-174
- Banerjee A., 1999, Panel data unit roots and cointegration : An overview, *Oxford Bulletin of Economics and Statistics*, Special issue, 61, pp. 607-629
- Banerjee A., M. Marcellino and C. Osbat, 2001, Testing for PPP : Should we use panel methods ? mimeo, European University Institute
- Basawa J.V. et alii, 1991, Bootstrapping unstable first-order autoregressive processes , *The annals of Statistics*, 19, pp. 1098-1101
- Becker B.J., 1997, P-values, combination of -, in S. Kotz, C.R. Read et D.L. Banks (edit), *Encyclopedia of Statistical Sciences*, update Vol. 1, Wiley, pp. 448-453
- Berkowitz J. and L. Killian, 1996, Recent developments in bootstrapping time series, mimeo
- Chang Y., 2000, Bootstrap unit root tests in panels with cross-sectional dependency, mimeo, Rice University
- Choi I., 2001, Unit root tests for panel data, *Journal of International Money and Finance*, 20, pp. 249-272
- Coakley J., A.M. Fuertes, 1997, New panel unit root tests of PPP, *Economic Letters*, 57, pp. 17-22
- Culver S.E., D.H. Papell, 1997, Is there a unit root in the inflation rate ? Evidence from sequential break and panel data models, *Journal of Applied Econometrics*, 12, pp. 435-444
- De Angelis D., S. Fachin, and C.A. Young, 1997, Bootstrapping unit root tests, *Applied Economics*, 29, pp. 1155-1161
- ESS, 1988, Tests and P-values, combining, in *Encyclopedia of Statistical Sciences*, S. Kotz et N. Johnson (edit), Vol. 9, Wiley, p. 216
- Fleissig A.R., J. Strauss, 1999, Is OECD real per capita GDP trend or difference stationary ? Evidence from panel unit root tests, *Journal of Macroeconomics*, 21, pp. 673-690
- Folks J.L., 1984, Combination of independence tests, in *Handbook of Statistics*, Vol. 4, P.R. Krishnaiah et P.K. Sen (eds.), Elsevier Science, pp. 113-121
- Heimonen K., 1999, Stationarity of the European real exchange rates evidence from panel data, *Applied Economics*, 31, pp. 673-677

Higgins M., E. Zakrajsek, 2000, Purchasing power parity : Three stakes through the heart of the unit root null, mimeo, Federal Bank of New York, April

Ims K.S., M.H. Pesaran and Y. Shin, 1997, Testing for unit roots in heterogeneous panels, WP, Department of Applied Economics, University of Cambridge

Johansen S., 1991, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregression models, *Econometrica*, 59, pp. 1551-1580

Karlsson S. et M. Löthgren, 2000, On the power and interpretation of panel unit root tests, *Economic Letters*, 66, pp. 249-255

Levin A., C.F. Lin, 1992, Unit root tests in panel data : asymptotic and finite sample properties, mimeo, University of California, San Diego

Lovell M.C., 1983, Data mining, *Review of Economics and Statistics*, 55, pp. 1-12

Lyhagen J., 2000, Why not use standard panel unit root test for testing PPP, Stockholm School of Economics, WP n° 413

Mac Donald R., 1996, Panel unit root tests and real exchange rates, *Economic Letters*, 50, pp. 7-11

Maddala G.S., 1999, On the use of panel data methods with cross country data, *Annales d'Economie et de Statistiques*, n° 55-56; pp. 429-449

Maddala G.S., S. Wu, 1999, A comparative study of panel data unit root tests and a simple alternative, *Oxford Bulletin of Economics and Statistics*, 61, pp. 631-652

O'Connell P.G., 1998, The overvaluation of purchasing power parity, *Journal of International Economics*, 44, pp. 1-19

Oh K.Y., 1996, Purchasing power parity and unit roots tests using panel data, *Journal of International Money and Finance*, 15, pp. 405-418

Papell D.H., 1997, Searching for stationarity : Purchasing power parity under the recent float, *Journal of International Economics*, 43, pp. 313-332

Psaradakis Z., 2000, P-value adjustments for multiple tests for non linearity, to appear in *Non Linear Dynamics and Econometrics*

Quah D., 1994, Exploiting cross section variation for unit root inference in dynamic data, *Economic Letters*, 44, pp. 9-19

Rao C.R., 1952, *Advanced statistical methods in biometric research*, Wiley



Sarantis N. and C. Stewart, 1999, Is the consumption ratio stationary ? Evidence from panel unit root tests, *Economic Letters*, 64, pp. 309-314

Taylor M.P. and L. Sarno, 1998, The Behavior of real exchange rates during the post Bretton Woods period, *Journal of International Economics*, 46, pp. 281-312

Wu J.L., 2000, Mean reversion of the current account : evidence from the panel data unit root tests, *Economic Letters*, 66, pp. 215-222

Wu Y., 1996, Are real exchange rates nonstationary ? Evidence from a panel data test, *Journal of Money, Credit and Banking*, 28, pp. 54-63

Wu Y., H. Zhang, 1996, Mean reversion in interest rates : New evidence from a panel of OECD countries, *Journal of Money, Credit and Banking*, 28, pp. 604-622

## APPENDIX : Figures and complementary tables

Figure 1 : Empirical percentage of rejection under alternative hypothesis :  
 $\frac{1}{2} = 0.8$ ;  $\alpha = 5\%$ ;  $N = 20$ ;  $T = 100$

Figure 2 : Distribution of estimated  $\hat{\alpha}$ , under  $\frac{1}{2} = 0.8$ ;  $N = 20$ ;  $T = 100$ ;  
IPS and DF statistics

Figure 3 : Distribution of estimated  $\hat{\alpha}$ , under  $\frac{1}{2} = 0.8$ ;  $N = 20$ ;  $T = 100$ ;  
 $\hat{\alpha}$  and marginal DF,  $t^m$ ; statistics

Table 7 : Empirical results, bilateral DM, monthly

Table 8 : Empirical results, bilateral DM, quarterly

table 9 : Empirical results, bilateral USD, quarterly

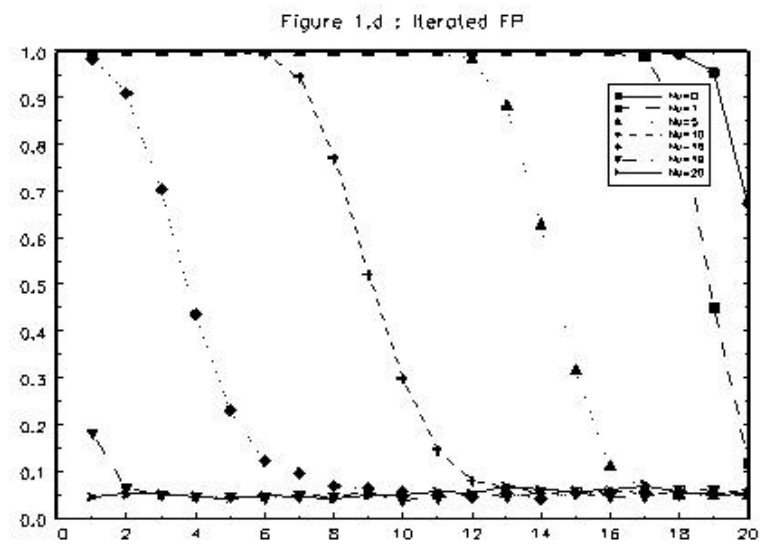
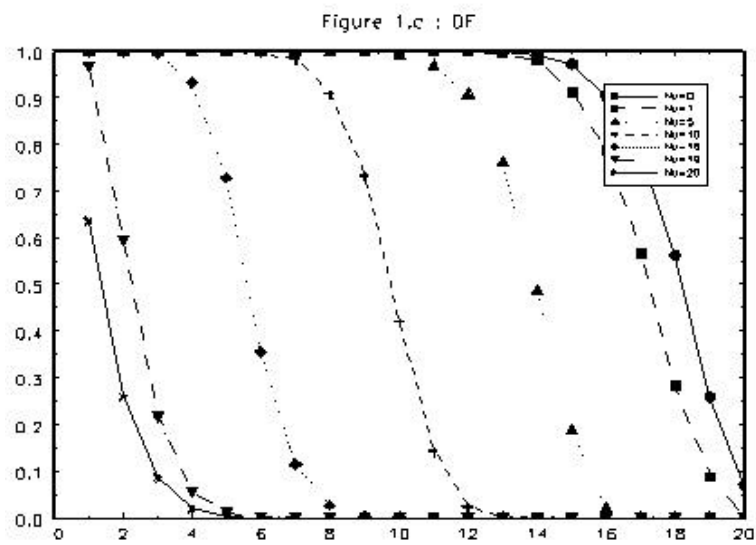
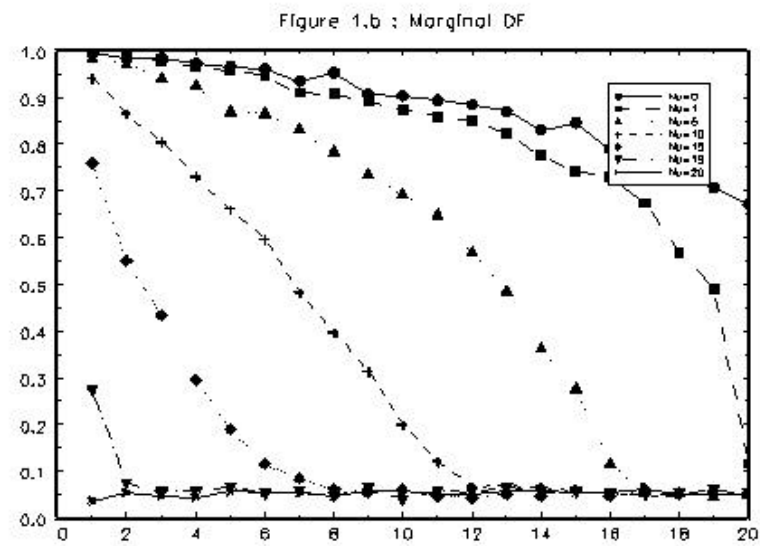
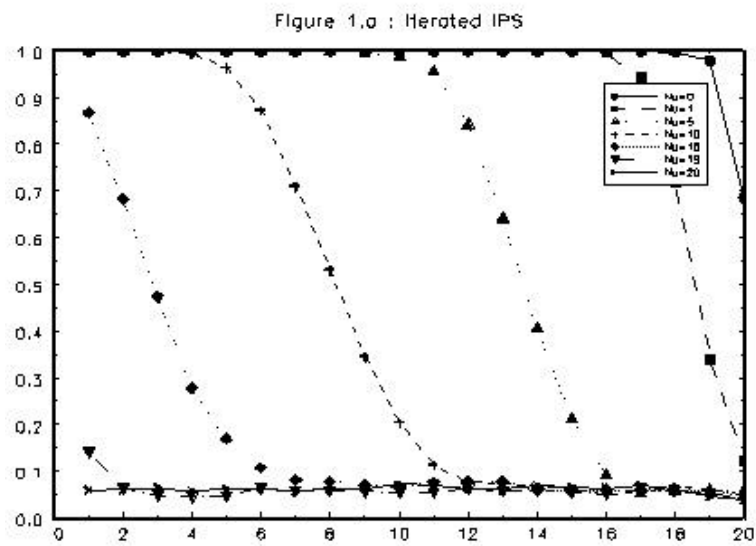


Figure 1: Empirical percentage of rejection under alternative hypothesis :  $\frac{1}{2} = 0.8$ ;  $\alpha = 5\%$ ;  $N = 20$ ;  $T = 100$

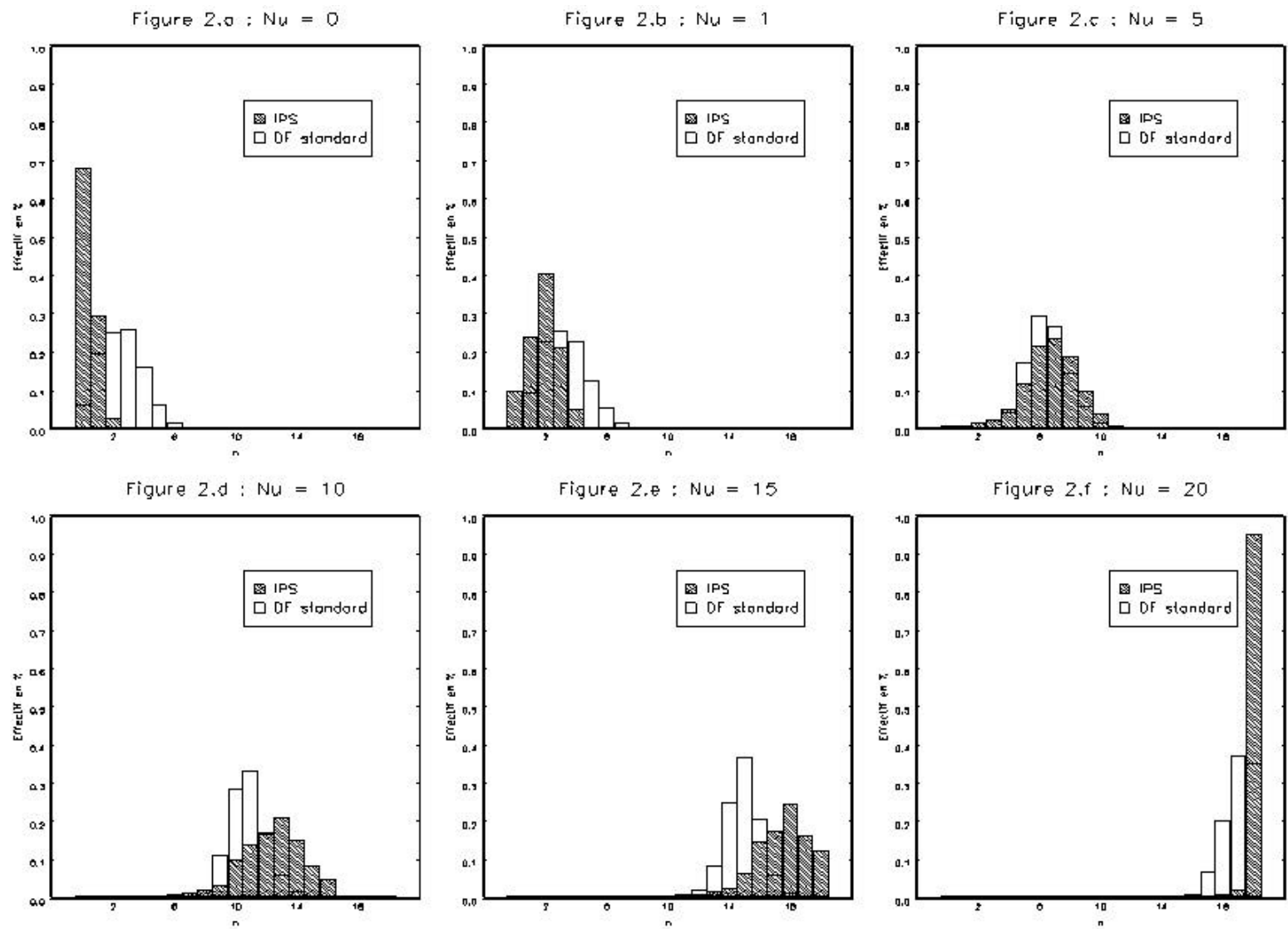


Figure 2: Distributed of estimated  $\hat{n}$ ; under  $\frac{1}{2} = 0.8$ ;  $N = 20$ ;  $T = 100$ ; IPS and DF statistics

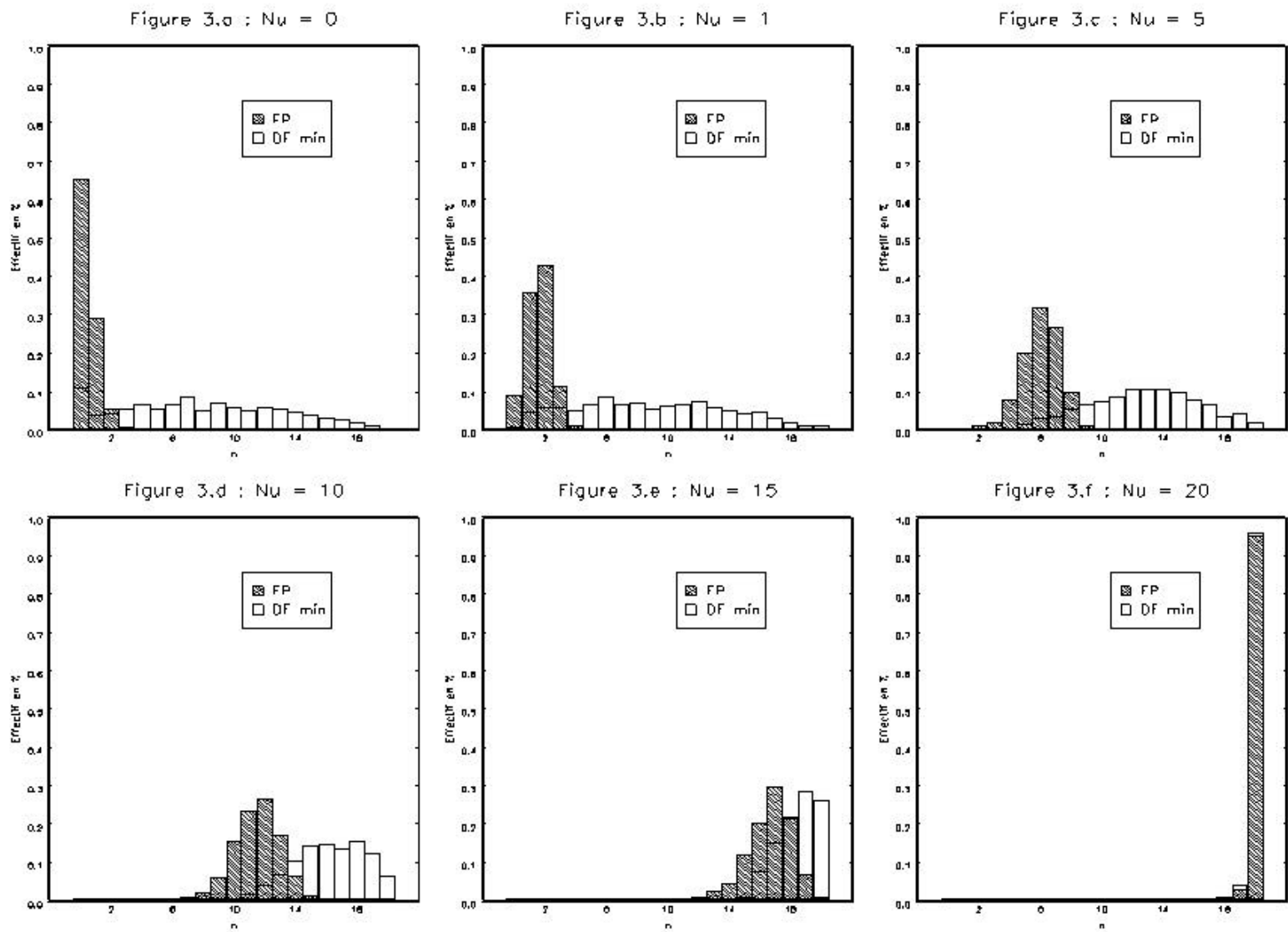


Figure 3: Distribution of estimated  $\hat{n}$ ; under  $\frac{1}{2} = 0.8$ ;  $N = 20$ ;  $T = 100$ ;  $P_{\lambda}$  and marginal DF,  $t^m$ ; statistics

Rank	Italy	Neth	Spain	Denm	UK	Fin	Belg	Switz	Port	Fr	Can	US	Swed	Norw	Austr	Jap
$\hat{t}_i$ (historical data)	-4.61 <sup>a</sup>	-4.43 <sup>a</sup>	-3.99 <sup>a</sup>	-3.94 <sup>a</sup>	-3.60 <sup>a</sup>	-3.47 <sup>a</sup>	-3.35 <sup>a</sup>	-3.24 <sup>a</sup>	-3.03 <sup>a</sup>	-2.5	-2.05	-1.91	-1.87	-1.75	-1.69	-1.06
$p(t_n)$	-	0	0	0	0	1	1	1	2	3	12	27	33	35	41	44
$\hat{\rho}_3$	116.67 <sup>a</sup>	101.67 <sup>a</sup>	84.11 <sup>a</sup>	75.21 <sup>a</sup>	59.99 <sup>a</sup>	53.14 <sup>a</sup>	43.42 <sup>a</sup>	30.04	24.64	16.60	14.08	16.32	16.06	9.93	7.07	1.27
CV $\hat{\rho}_3; n; p(t_n)$	45.84	43.23	41.04	38.16	36.51	33.84	31.65	31.33	28.08	26.16	21.47	18.32	16.61	13.63	9.89	5.73
$\hat{\beta}_3$	-6.48 <sup>a</sup>	-5.75 <sup>a</sup>	-5.09 <sup>a</sup>	-4.48 <sup>a</sup>	-3.93 <sup>a</sup>	-3.40 <sup>a</sup>	-2.87 <sup>a</sup>	-2.34 <sup>a</sup>	-1.83 <sup>a</sup>	-1.78 <sup>a</sup>	-2.10 <sup>a</sup>	-2.03 <sup>a</sup>	-1.70 <sup>a</sup>	-1.46 <sup>a</sup>	-1.00	-1.17
CV $\hat{\beta}_3; n; p(t_n)$	-2.21	-2.12	-2.15	-2.09	-1.81	-1.81	-1.77	-1.87	-1.65	-1.59	-1.63	-1.51	-1.45	-1.32	-1.19	-0.99

Table 7 : Empirical results, bilateral DM, monthly

Rank	Italy	Spain	Neth	UK	Denm	Fin	Switz	Port	Fr	Belg	Can	US	Austr	Norw	Swed	Jap
$\hat{t}_i$ (historical data)	-4.35 <sup>a</sup>	-3.84 <sup>a</sup>	-3.66 <sup>a</sup>	-3.56 <sup>a</sup>	-3.49 <sup>a</sup>	-3.15 <sup>a</sup>	-2.98 <sup>a</sup>	-2.79	-2.64	-2.14	-1.92	-1.75	-1.67	-1.67	-1.51	-1.19
$p(t_n)$	-	0	0	1	1	1	3	4	6	9	23	32	41	45	45	53
$\hat{\rho}_3$	89.55 <sup>a</sup>	77.00 <sup>a</sup>	68.80 <sup>a</sup>	59.95 <sup>a</sup>	50.46 <sup>a</sup>	38.42	33.45	26.59	21.02	14.60	15.29	14.52	15.64	16.62	5.70	2.27
CV $\hat{\rho}_3; n; p(t_n)$	48.83	46.64	46.16	49.73	45.36	39.12	35.23	34.27	31.39	26.48	23.07	19.97	16.61	14.05	9.63	6.69
$\hat{\beta}_3$	-5.20 <sup>a</sup>	-4.61 <sup>a</sup>	-4.09 <sup>a</sup>	-3.59 <sup>a</sup>	-3.06 <sup>a</sup>	-2.69 <sup>a</sup>	-2.29	-1.96	-1.62	-1.96	-2.05	-2.14	-1.96	-1.55	-1.32	-1.19
CV $\hat{\beta}_3; n; p(t_n)$	-2.67	-2.79	-2.81	-3.00	-2.95	-2.63	-2.37	-2.24	-2.09	-1.49	-2.03	-1.88	-1.68	-1.60	-1.39	-1.04

Table 8 : Empirical results, bilateral DM, quarterly

Rank	Port	UK	Italy	Switz	Fin	Spain	Fr	Denm	Belg	Austr	Ger	Can	Norw	Neth	Swed	Jap
$\hat{t}_i$ (historical data)	-2.73	-2.70	-2.53	-2.28	-2.24	-2.18	-1.84	-1.70	-1.64	-1.57	-1.56	-1.51	-1.33	-1.26	-1.24	-0.61
$p(t_n)$	-	7	8	11	18	19	22	36	43	47	50	51	53	62	65	66
$\hat{\rho}_3$	36.56	42.38	32.93	28.93	30.90	24.56	18.09	22.17	24.16	22.28	22.48	14.48	9.31	10.41	9.46	1.01
CV $\hat{\rho}_3; n; p(t_n)$	64.65	66.87	65.87	59.00	59.67	54.60	49.40	42.80	38.07	33.33	29.47	25.07	20.80	16.20	11.40	7.33
$\hat{\beta}_3$	-2.12	-1.79	-1.67	-1.79	-1.55	-1.32	-1.94	-2.14	-2.07	-1.99	-1.69	-1.45	-1.59	-1.35	-0.83	-1.17
CV $\hat{\beta}_3; n; p(t_n)$	-4.13	-4.01	-4.07	-3.48	-3.65	-3.39	-3.28	-2.92	-2.67	-2.47	-2.27	-2.07	-1.89	-1.64	-1.37	-1.05

Table 9 : Empirical results, bilateral USD, quarterly