An Improved Panel Unit Root Test Using GLS-Detrending

Claude Lopez[1]
University of Cincinnati
August 2004

This paper offers a panel extension of the unit root test proposed by Elliott, Rothenberg and Stock (1996). More specifically, the proposed approach allows for heterogeneous serial and contemporaneous correlation, while fixing the rate of convergence to be homogeneous across series. The new test demonstrates significantly better finite sample-power properties than the Levin, Lin and Chu (2002) or the Im, Pesaran and Shin (2003) tests, especially for highly persistent series. An application to the real exchange rate convergence illustrates the impact of such improvements. Analysing the post Bretton Woods period, the new test provides strong and reliable evidence of Purchasing Power Parity.

JEL classification: C12, C15, C32, C33

[1]I am very grateful to Chris Murray and David Papell for their encouragements and helpful discussions. I would like to thank Debashis Pal, two referees and the participants in seminars at the University of Houston and at the 11th International Conference on Panel Data for helpful comments and discussions. Claude Lopez, Assistant Professor, Department of Economics - University of Cincinnati, 1209 Crosley Tower Cincinnati OH 45221-0371, Tel: 513-556-2346, Fax: 513-556-2669, Email: Claude.Lopez@uc.edu. 513-556-2669, Email: Claude.Lopez@uc.edu.

1 Introduction

Economic analysis of most time series requires stationarity of the data. Unit root tests are commonly used to address this matter. The most well-known among them is the augmented Dickey-Fuller (ADF) unit root test. Recent works, however, have acknowledged the poor power properties of this test, which leads to a vast literature attempting to overcome these disadvantages. These developments have occurred at both panel and univariate levels.

Authors such as Levin, Lin and Chu (LLC) (2002), Im, Pesaran and Shin (2003) and Maddala and Wu (1999) offer excellent alternatives to the ADF test by combining time-series information with cross-sectional variability. The panel approach appears extremely appealing for two reasons. First, the inclusion of a limited amount of cross-sectional information induces significant improvements in term of power. Second, the data needed for this type of analysis is increasingly available.

The LLC test and more specifically the LLC hypotheses are widely used. Indeed, several works propose enhanced versions of this test, producing data specific estimations. Papell (1997) suggests accounting for heterogeneous serial correlation, while O'Connell (1998) demonstrates the necessity of allowing for the cross-sectional dependence in the estimation procedure.[2,3] Papell and Theodoridis (2001) incorporate both by considering a panel version of the ADF test, using the LLC hypotheses and allowing for heterogeneous serial and contemporaneous correlation.

At the univariate level, Elliott, Rothenberg and Stock (1996) develop a GLS-detrended/demeaned version of the ADF test. Running the ADF test on the GLS-transformed data leads to one of the most powerful univariate test: the DF-GLS test.[4] While these innovations deliver substantial gains in power

[2] Papell (1997) shows a strong relation between the size of the panel and the rate of rejection of the unit root hypothesis.

[3] O'Connell (1998) points out the sizeable bias induced by the neglect of contemporaneous correlation when estimating cross-correlated data.

[4] Hansen (1995) proposes a more powerful alternative to the DF-GLS test by including

over the univariate ADF test, they still demonstrate limited performance when applied to economic time-series data. Indeed, most of the data sets available have a limited length.

The present work intends to develop a new estimation procedure, offering satisfying performance especially in cases of highly persistent series and limited amount of data. Seeking a significant increase in power over existing tests, we combine the GLS-detrending of Elliott, Rothenberg and Stock (1996) with the panel ADF test, using Levin, Lin and Chu's (2002) hypotheses. We analyze the behavior of the new approach for various sample sizes, panel widths, and degrees of persistence with a Monte Carlo experiment. The main result is that the new panel unit root test displays significantly better finite sample power than existing univariate and panel unit root tests.

We illustrate the impact of such improvements by analyzing the Purchasing Power Parity (PPP) query within industrialized countries. We focus on the post Bretton-Woods period because neither the panel ADF test nor the DF-GLS test are able to reject the existence of a unit root. The principal outcome is a robust overall support for the PPP hypothesis, independently of the width and the length of the panels considered.

The next section provides a concise review of the literature that relates to the understanding of the proposed unit root test. Section 3 develops the new testing procedure and conducts detailed size and power experiments. Section 4 presents an empirical application to PPP, and, finally, Section 5 summarizes our findings.

2 Existing Unit Root Tests: A Concise Review of the Literature

In this section, we present the existing tests which have motivated this work and which help to its understanding. First, the standard ADF unit root test

covariates to the test, at the univariate level.

runs the following regression:

$$y_t = d_t + \alpha y_{t-1} + \sum_{i=1}^{k} \psi_i \Delta y_{t-i} + u_t \tag{1}$$

where y_t is the tested series, d_t a set of deterministic regressors, k the lagged first difference terms allowing for serial correlation and u_t the error term of the regression. The unit root null hypothesis is that $\alpha = 1$, and the alternative of stationarity is $\alpha < 1$. This test is well-known for its poor power, and the subsequent literature suggests several alternatives. In the next two subsections, we describe some recent developments at both panel and univariate levels.

2.1 More Powerful Unit Root Tests: Panel ADF Tests

The idea behind the panel unit root tests is to combine cross-sectional and time-series information to achieve a more efficient test. Several panel procedures have been developed to test the unit root null hypothesis against various alternative hypotheses. Levin, Lin, and Chu (2002) test the unit root null hypothesis against the homogenous alternative that *every* series in the panel is stationary with the same speed of reversion. Im, Pesaran and Shin (2003) and Maddala and Wu (1999) test the unit root null hypothesis against the alternative that *at least one* series in the panel is stationary. The new test, later proposed in this paper, focuses on the stationarity of the *entire* panel, which automatically leads us to concentrate on the LLC framework. The LLC test runs the following panel version of equation (1):

$$y_{jt} = d_{jt} + \alpha y_{j,t-1} + \sum_{i=1}^{k_j} \psi_{ji} \Delta y_{j,t-i} + u_{jt} \tag{2}$$

where α is the homogeneous rate of convergence of the panel. The null hypothesis is that $\alpha = 1$ and the alternative is that $\alpha < 1$. For each series

j, $j = 1, ..., N$, $d_{it} = \beta_j' z_t$ is a set of deterministic regressors, which allows for heterogeneous intercepts and time trends and k_j lagged first differences term are included to account for serial correlation. The error terms are assumed to be contemporaneously uncorrelated, $E(u_{it} u_{jt}) = 0$ for $i \neq j$.

While the LLC test leads to substantial improvements over the ADF test in terms of power, it is based on the extremely restrictive assumption that the series in the panel are cross-sectionally uncorrelated. Maddala and Wu (1999) and O'Connell (1998) demonstrate that if the error terms in equation (2) are indeed contemporaneously correlated, the LLC test exhibits severe size distortions.[5] As an alternative, Papell and Theodoridis (2001) estimate the system of equations defined by (2) using Seemingly Unrelated Regressions (SUR). This version of the LLC test, named the ADF-SUR test, accounts for serial and contemporaneous correlation. In the rest of this paper, we shall estimate equation (2) allowing for contemporaneous correlation.

Performing the ADF-SUR test is a two-step procedure. First, for each series j, $j = 1, ..., N$, the number of lagged first difference terms, k_j, must be selected to account for serial correlation. In this work, we use the recursive lag-selection procedure of Hall (1994) and Ng and Perron (1995). Then, having selected k_j, the system of equations needs to be estimated via SUR, constraining the values of α to be identical across equations.

2.2 More Powerful Univariate Unit Root Tests: The DF-GLS Test

Elliott, Rothenberg, and Stock (1996) construct an efficient univariate unit root test based on local-to-unity asymptotic theory. The DF-GLS test is an ADF test on GLS-demeaned (or GLS-detrended) data. Specifically, the DF-GLS test runs the following regression:

[5] However, by imposing homogeneous serial correlation properties across the series, O'Connell (1998) tends to under reject the null hypothesis.

$$y_t^d = \alpha y_{t-1}^d + \sum_{i=1}^{k} \psi_i \Delta y_{t-i}^d + u_t \quad (3)$$

where y_t^d is the transformed data such that $y_t^d = y_t - \tilde{\beta} z_t$. $\tilde{\beta}$ is the least-squares estimate of the regression of $\tilde{z}$ on $\tilde{y}$. $\tilde{y}_t$ and $\tilde{z}_t$ are the quasi-differences of y_t and z_t respectively, i.e. $\tilde{y}_t = (y_1, y_2 - ay_1, ..., y_T - ay_{T-1})'$, and $\tilde{z}_t = (z_1, z_2 - az_1, ..., z_T - az_{T-1})'$. $a = 1 + \frac{\bar{c}}{T}$ represents the local alternative, with $\bar{c} = -7$ when $z_t = (1)$ and $\bar{c} = -13.5$ when $z_t = (1, t)'$.[6] The standard hypotheses are tested: $H_0 : \alpha = 1$ versus $H_1 : \alpha < 1$.

The lag-selection issue in the DF-GLS regressions has received much attention recently. Ng and Perron (2001) propose a new lag selection procedure, the Modified Akaike Information Criterion (MAIC), that provides the best combination of size and power in finite samples when combined with the GLS-transformation.[7] In subsequent applications, we employ the MAIC when performing the DF-GLS test.

3 An Improved Panel Unit Root Test : The DF-GLS-SUR Test

Both the ADF-SUR and the DF-GLS tests demonstrate higher power than the standard ADF test. They display, however, a limited ability to reject the unit root hypothesis for economic time series of the length generally encountered in practice. Consequently, we propose to combine both innovations to obtain a more powerful unit root test. The new test, named the DF-GLS-

[6] ERS (1996) show that $\bar{c}$ = -7 ($\bar{c}$= -13.5) corresponds to the tangency between the asymptotic local power function of the test and the power envelope at 50% power in the case with a constant (the case with a constant and a trend).

[7] MAIC takes into account the nature of the deterministic components and the demeaning/detrending procedure, which allows a better measurement of the cost of each lag-length choice.

SUR test, runs the following system of equations for $j = 1, ..., N$:

$$y_{jt}^d = \alpha y_{j,t-1}^d + \sum_{i=1}^{k_j} \psi_i \Delta y_{j,t-i}^d + u_{jt} \tag{4}$$

where α is the homogeneous rate of convergence of the panel. The standard hypotheses are tested, that is $H_0 : \alpha = 1$ versus $H_1 : \alpha < 1$.

The DF-GLS-SUR test requires a three-steps procedure. For each series j, the data needs first to be GLS-transformed, then k_j, the number of lagged first difference terms allowing for serial correlation, must be selected using MAIC. Finally, the system of equations is estimated via SUR, constraining the values of α to be equal across equations and using the pre-selected k_j. This procedure allows to test for the stationarity of the entire panel while accounting for data specific serial and contemporaneous correlation.

3.1 Finite Sample Performances: Power Analysis

For the remaining of the paper, we consider the ADF-SUR test as benchmark. The Monte Carlo experiment considers panels with $T \in \{25, 50, 75, 100, 125, 250\}$ and $N \in \{5, 10, 15, 20\}$.[8] The data generating processes follow:

$$y_{jt} = \rho y_{j,t-1} + u_{jt} \tag{5}$$

where $u_{jt}\,\tilde{}\, iidN(0, 1)$, $E(u_{it}u_{jt}) = 0$ for $i \neq j$. We consider the following alternatives $\rho \in \{1, 0.99, 0.97, 0.95, 0.90, 0.85, 0.80\}$ where $\rho = 1$ allows us to generate finite sample critical values, while $\rho < 1$ are the alternative considered for the finite sample power simulations. *Tables* 1 and 2 display the level of power for the ADF-SUR and the DF-GLS-SUR tests when studying AR(1) processes, respectively with a constant and with a constant and a

[8] We consider $T \in \{35, 50, 75, 100, 125, 250\}$ for the case with heterogeneous constants and trends.

trend.[9,10] The nominal size considered is 5%.

Below, we discuss three aspects of the results for both tests: a change in T, the length of the panel, a change in N, the width of the panel, and a decrease in ρ, the persistence of the series.

3.1.1 Demeaned case: *Table 1*

First of all, a comparison between the power of the DF-GLS-SUR test and of the DF-GLS test demonstrates that the inclusion of few more series to the univariate DF-GLS test leads to drastic power improvements: for $\rho = 0.95$ and $T = 100$, the power of the DF-GLS test is equal to 0.26, while for the DF-GLS-SUR test the power is 0.998, with $N = 5$.

Then, $Table$ 1 shows that a sole increase in T leads to consistent increases in power for both tests with significantly stronger improvements for the DF-GLS-SUR test than the ADF-SUR test. For example, a highly persistent system of series, $\rho = 0.99$, a limited amount of series, $N = 5$, and a small increase in the length of the panel, T varies from 25 to 50 observations, the DF-GLS-SUR test produces an increase in power ten times higher than the ADF-SUR test. The same case with a wider panel ($N = 20$) demonstrates a similar outcome. For highly persistent series with a small number of observations, the DF-GLS-SUR test offers higher power than the ADF-SUR test, and takes better advantage of an increase in T.

For less persistent processes, for example $\rho \in \{0.97, 0.95\}$, the DF-GLS-SUR test still presents a stronger response to an increase in T than the ADF-SUR test.

[9] We also considered the AR(p), p>1, case and found, as expected that the lag selection induces a uniform power loss, compared to the case where the lag length is known and equal to 0. However, the AR(1) and AR(p) cases present similar patterns: the DF-GLS-SUR test demonstrates an overall higher power than the ADF-SUR test. Therefore, we just present the AR(1) results. The other results are available upon request.

[10] $Tables$ 1 and 2 report only the size-adjusted power of both tests for $\rho \in \{1, 0.99, 0.97, 0.95, 0.90\}$ and $T \in \{25, 50, 75, 100, 125\}$. The results for $\rho \in \{0.85, 0.80\}$ and $T = 250$ are available upon request.

A sole increase in N leads to consistent power improvements, with a stronger impact on the DF-GLS-SUR test than on the ADF-SUR test. For example, with $\rho = 0.99$, and $T = 25$, an increase in the number of series, N evolving from 5 to 10, leads to a power increase for the DF-GLS-SUR test twelve times stronger than for the ADF-SUR test. Furthermore, if we compare the impact of a change in T with the impact of a change in N, the results show for both tests that an increase in N has a stronger effect than an increase in T. Considering $\rho = 0.95$, the DF-GLS-SUR test has a power of 0.83 for $(N, T) = (5, 50)$. A minimum of 50 observations needs to be added to reach a power level close to 1, while the addition of only 5 series displays the similar result.

$(N, T) = (20, 50)$ presents an interesting case, especially if the processes are highly persistent. The ADF-SUR test is well-known for its power deficiency when the width and the length of the panel are too close: with $\rho = 0.99$, this combination presents a power of 0.56 for the DF-GLS-SUR test and of 0.09 for the ADF-SUR test. If $\rho = 0.97$, the DF-GLS-SUR test reaches a power of 0.99 while the ADF-SUR test offers only a power of 0.19. The DF-GLS-SUR test demonstrates an impressive higher power than the ADF-SUR test for panels including highly persistent series and with N close to T.

As expected, a change in the series persistence also has a major influence on the behavior of both the DF-GLS-SUR and the ADF-SUR tests. For $(N, T) = (10, 75)$, the DF-GLS-SUR test presents a power of 0.50 for $\rho = 0.99$, of 0.99 for $\rho = 0.97$ and of 1.00 for $\rho = 0.95$. More generally, the DF-GLS-SUR test has a power of 1.00 or close to 1.00 whenever $\rho \leq 0.97$ and $T \geq 75$.

3.1.2 Detrended case: *Table 2*

The addition of a trend to the regressions leads to a uniform loss in power for both tests: with $\rho = 0.97$ and $(N, T) = (15, 100)$, the DF-GLS-SUR test produces a power of 1.00 in the demeaning case but reaches only a power of 0.30 in the detrending case. However, combining time-series information with cross-sectional information still provides significant improvements in the test performance. For $\rho = 0.95$ and $T = 100$, the DF-GLS$^\tau$ test reaches a power level of 0.10 while the DF-GLS-SUR$^\tau$ test, accounting for four more series $(N = 5)$, is able to achieve a power level of 0.33.

The GLS-transformation shows a similar impact on the size-adjusted power than in *Section 3.1.1.* If $\rho = 0.95$ and $(N, T) = (5, 125)$, the DF-GLS-SUR test has a power of 0.47 while the ADF-SUR test offers a power level of 0.34.

Furthermore, the amplitude of these enhancements varies following changes in T, in N or in ρ, as well as the test considered. For $\rho = 0.97$ and $N = 20$, a raise in T from 50 to 100 observations induces an power increase of 0.25 for the DF-GLS-SUR test and of 0.13 for the ADF-SUR test. Likewise, if N varies from 10 to 20 series, when $\rho = 0.97$ and $T = 100$, the power augments by 0.15 for the DF-GLS-SUR test and by 0.06 for the ADF-SUR test. Finally, a decrease in the persistence, from $\rho = 0.97$ to $\rho = 0.95$, generates strong improvements in the performance for both tests: for $(N, T) = (10, 125)$, the observed power increase is of 0.48 for the DF-GLS-SUR test and of 0.37 for the ADF-SUR test.

To sum up, this analysis reveals two major outcomes. First, as expected, by incorporating cross-sectional variations we are able to significantly enhance the power of the univariate DF-GLS test. Secondly, the comparison of both panel-test performances demonstrates strong improvements due to the GLS-transformation. Overall, the DF-GLS-SUR test has a higher finite sample power than the ADF-SUR test. Furthermore, each increase in information (either N or T) results in larger power increases for the DF-GLS-SUR test than for the ADF-SUR test. Finally, the DF-GLS-SUR test presents an attractively high power for small panels and for highly persistent series.

3.2 Robustness Analysis: Figures 1 and 2

One obvious objection to this new test stands in the homogeneity imposed by the alternative hypothesis. This restriction seems to undermine the enhanced power properties previously presented. Consequently, in this section, we focus on the impact of this constraint by measuring the performance of the DF-GLS-SUR test when applied to series with heterogeneous rates of convergence. Our results show that, even under such conditions, the DF-GLS-SUR test remains one of the most powerful unit root tests available.

We proceed with the Monte Carlo experiment defined earlier for the power analysis, but allowing for the rate of convergence to vary across the generated series, i.e. the data generating process follows:

$$y_{jt} = \rho_j y_{j,t-1} + u_{jt} \tag{6}$$

where u_{jt}~$iidN(0,1)$, $E(u_{it}u_{jt}) = 0$ for $i \neq j$,and $\rho_j \leq 1$. Again, we focus on estimating AR(1) equations in the demeaning case ($z_t = 1$).[11]

Due to the infinite number of cases existing, we focus on the six subsequent panels:

$N = 5$, $T \in \{50, 100\}$ and ρ_j are divided in two groups such that $\{\rho_i, \rho_l\} \in \{1.00, 0.99, 0.97, 0.95, 0.90, 0.85, 0.80\}$ for $i \in \{1, 2, 3\}$ and $l \in \{4, 5\}$.

$N = 15$, $T \in \{50, 100\}$ and ρ_j are divided in three groups such that $\{\rho_i, \rho_l\} \in \{1.00, 0.99, 0.97, 0.95, 0.90, 0.85, 0.80\}$ for $i \in \{1, 2, 3, 4, 5\}$, and $l \in \{6, 7, 8, 9, 10\}$ and $\rho_m \in \{0.80, 0.95\}$ for $m \in \{11, 12, 13, 14, 15\}$.

The size-adjusted power resulting from these simulations is reported in $Figure$ 1.

The 3D graphs demonstrate strong deteriorations of the DF-GLS-SUR-test performance in presence of random walks among the series. For example, if $(N, T) = (5, 50)$ and $(\rho_i, \rho_l) = (1.00, 0.99)$, the power level reached is 0.08

[11] This allows us to compare the size-adjusted power with Bowman (1999) and Im, Pesaran and Shin (1997).

instead of 0.18 when $(\rho_i, \rho_l) = (0.99, 0.99)$. Power losses are also observed when some of the series estimated include processes more persistent than the alternative considered in the homogeneous case: if $(N, T) = (15, 50)$, the DF-GLS-SUR test achieves a power of 0.88 when $(\rho_i, \rho_l, \rho_m) = (0.97{,}0.99{,}0.80)$, instead of 1.00 when $(\rho_i, \rho_l, \rho_m) = (0.80, 0.80, 0.80)$. The converse is also verified: if $(\rho_i, \rho_l, \rho_m) = (0.85, 0.90, 0.95)$, the power equals 1.00, while if $(\rho_i, \rho_l, \rho_m) = (0.95, 0.95, 0.95)$, it equals 0.99.

The relatively poor performance of the DF-GLS-SUR test in presence of non-stationary processes encourages a comparison with the Im, Pesaran and Shin (IPS) (2003) test. $Figure$ 2 graphs the power simulations such that the X-axis represents the number of stationary series in the panel and the Y-axis is the power. The panels considered are such that $N \in \{5, 10, 15, 20\}$ and $T = 100$. The rates of convergence for the stationary processes are $\rho \in \{0.90, 0.95\}$.

If we consider $(N, T) = (10, 50)$, with 5 stationary processes converging at $\rho = 0.95$, the DF-GLS-SUR test has a power of 35% while for the IPS test it is of 25%. On the contrary, the same case with $\rho = 0.90$ reveals a power of 60% for the IPS test and of 40% for the DF-GLS-SUR. Overall, the DF-GLS-SUR test demonstrates a higher power than the IPS test when $\rho = 0.95$. These results imply that, in highly persistent cases, the impact of GLS-transformation prevails over the negative effect of the homogeneous alternative on the test performance. Our findings confirm that the GLS-transformation improves the finite sample-power properties of the test, especially when investigating mixes of highly persistent and non-stationary series, even though the alternative hypothesis is wrong.

To sum up, the DF-GLS-SUR test alternative hypothesis has a limited impact on the test performance in presence of series converging at different rates. The DF-GLS-SUR test was designed to answer more accurately whether or not the panel converges. Its overall satisfying performance in presence of homogeneous or heterogeneous rates of convergence in stationary data sets confirms its accuracy. Furthermore, the relatively low power achieved in presence of random walks in the panel is not a major issue because it is still significantly higher than the nominal size (5%).

3.3 Empirical Size

Finally, we need to measure the impact of the cross-sectional dependency on the test size. We first focus on AR(1) processes to identify the specific effect of contemporaneous correlation. Then, we analyze more general AR(p) cases allowing for heterogenous serial and contemporaneous correlation.

To analyze the size in AR(1) cases, we use the finite sample critical values generated in *section 3.1* and consider different degrees of cross-sectional dependence of the form:

$$\Omega = \begin{bmatrix} 1 & \omega & ... & \omega \\ \omega & 1 & ... & \omega \\ . & . & . & . \\ \omega & \omega & ... & 1 \end{bmatrix}$$

with $\omega \in \{0.25, 0.50, 0.75\}$ with $E(u_{i,t}, u_{j,s}) = 0 \ \forall i, j$ if $s \neq t$.

Tables 3 and 4 present the empirical size at 5% significance level for the DF-GLS-SUR test for both demeaned and detrended cases. Both tables present a same pattern. In the case of weak dependence ($\omega = 0.25$), the test has a good size for all (N, T). However, size distortions appear while the degree of dependence increases, especially when N and T are relatively close and for T relatively small: for $(N, T) = (20, 25)$, the test has a size of 6.9% when $\omega = 0.25$, of 14% when $\omega = 0.50$ and of 26%, when $\omega = 0.75$. This issue disappears when T increases.

More general cases include serial and contemporaneous correlation. In order to measure the test-size distortion in presence of both correlations, we look at real data sets, testing for real exchange rate convergence within industrialized countries. The data and its analysis are presented next section. The focus here stays on the behavior of the test when applied to different panels with heterogenous serial and contemporaneous correlation. The bootstrap method is used to estimate empirical distributions of the test statistics. We then compare with critical values generated following the method described in *Section 3.1*.

In order to bootstrap the data, the non-diagonal variance-covariance matrix of the innovations is estimated as follow. The j^{th}series, $j = 1, ..., N$, are defined by:

$$y_{jt} = d_{jt} + \rho_j y_{j,t-1} + u_{jt} \tag{7}$$

where $u_{jt} = \beta_j u_{j,t-1} + \xi_{jt}$ with $(\xi_{1t}...\xi_{Nt})' \sim N(0_N, \Omega)$ and $E(u_{it}u_{jt}) \neq 0$ for $i \neq j$. The data generating process of ξ_{jt} is estimated then the u_{jt} and Σ, the variance-covariance matrix of the innovations, i.e. $(u_{1t}...u_{nt}) \sim N(0_N, \Sigma)$, are deduced. A total of 5000 replications are used to generate the empirical distributions. The resulting size distortions are reported $Table$ 5. Similar to the AR(1) case, the size distortions are stronger when N and T are close and when T is relatively small.

Both experiments conclude that the DF-GLS-SUR test show some size distortions in presence of cross-sectional correlation. The degree of correlation worsen this issue. Hence, we advocate to generate data specific critical values.

4 Illustration: Purchasing Power Parity

The real exchange rate convergence is a fundamental assumption in international economics. However, finding empirical evidence of Purchasing Power Parity for the post Bretton-Woods period remains a challenge. The test lack of power is commonly accepted as the explanation. Our results confirm that with an adequate tool, the DF-GLS-SUR test, we are able to provide strong evidence of PPP where the ADF, the DF-GLS and the ADF-SUR tests failed.

We consider quarterly CPIs and nominal exchange rates in dollars, from 1973(1), first quarter, to 1998(2), second quarter, (Source IFS, CD-Rom for 03/2002), for 21 industrialized countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Netherlands, Norway, New Zealand, Portugal, Spain, Sweden, Switzerland, the U.K., and the U.S..We then construct the corresponding real exchange rate, q_j (in logarithm) follows:

$$q_j = e_j + p^* - p_j \tag{8}$$

where e_j, p_j and p^* are the logarithm of the nominal exchange rate (U.S. dollar as numeraire), the foreign CPI and the US CPI.

We first proceed with univariate estimations of the real exchange rates through the ADF and the DF-GLS tests, using as lag-length selection the recursive and the MAIC procedures respectively. The results are shown in *Table* 6. Few rejections of the unit root hypothesis are observed: the ADF test rarely rejects while the DF-GLS test offers some rejections, varying from a 10% level for Denmark and Italy to a 5% level for Belgium, France, Germany, Greece and the Netherlands.

Next, we estimate the real exchange rates with the ADF-SUR and the DF-SUR-GLS tests. As advised previously, we generate data specific critical values via the bootstrap method presented *section* 3.3 .

The data is grouped such that the panel of the 20 U.S.-real exchange rates (All20) includes Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Netherlands, Norway, New Zealand, Portugal, Spain, Sweden, Switzerland, and the U. K. Then we consider the following panels: the European Community (EC), the European Monetary System (EMS), the 6 and 10 most industrialized countries (G6, G10), the Euro area as of 1999 (E10), the Euro area as of 2001 (E11), and the OECD countries (13).[12] For each panel, *Table* 7 reports the estimated α, and the corresponding t-statistic.

The panels considered vary in size with a width including between 6 and 20 US-real exchange rates and a length of 102 observations. As shown in the performance analysis, the DF-GLS-SUR test demonstrates an high power for these specific cases (a minimum power level of 90%) while the ADF-SUR test behaves poorly, at least for the small panels (a power level of 20%). For the studied panels, the bias of the DF-GLS-SUR test is negligible compared to the bias of the ADF-SUR test. Furthermore, the high power observed for the DF-GLS-SUR test combined with a size fixed at 5% implies that the results strongly reflect the information available in the data.

The DF-GLS-SUR test demonstrates uniformly stronger rejections, with 7 rejections at 1% and 1 at 5% while the ADF-SUR test shows a majority of rejections at 5% or less. By using a more powerful alternative to the existing tests, we are able to produce the strongest evidence of PPP for the floating period.

5 Conclusion

The literature already provides several more powerful alternatives to the ADF unit root test. However, all of them demonstrate a limited ability to reject correctly the unit root hypothesis when applied to highly persistent time series with a limited span.

This paper attempts to produce a more efficient testing procedure allow-

[12] EC includes Belgium, Denmark, France, Germany, Greece, Ireland, Italy, the Netherlands, Portugal, Spain, and the U.K. EMS includes Belgium, Denmark, France, Germany, Ireland, Italy, and the Netherlands. G6 includes Canada, France, Germany, Italy, Japan, and the U.K. For G10, Belgium, the Netherlands, Sweden, and Switzerland are added. E11 includes Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Portugal, and Spain. E10 does not include Greece. 13 includes Australia, Belgium, Canada, Denmark, Finland, France, Germany, Italy, the Netherlands, Norway, Sweden, and the U.K.

ing the reliable analysis of such data sets. The DF-GLS-SUR test is an extension of Elliott, Rothenberg, and Stock's (1996) GLS-transformation to a version of the Levin, Lin and Chu's (2002) test. Via Monte Carlo experiments, we show the interesting behavior of this new test. For both the demeaned and detrended cases, the DF-GLS-SUR test offers a uniformly higher finite-sample power than the ADF-SUR test. Furthermore, the DF-GLS-SUR-test performance remains attractive when studying data with heterogeneous rates of convergence across the series. Some size distortions appear when the data is highly cross-correlated so we advise to generate data specific critical values.

The most pertinent feature of the DF-GLS-SUR test stands in its satisfying power when applied to highly persistent processes with limited amount of observations. Indeed, it is always a challenge to increase significantly the time-series dimension of economic data while the cross-sectional dimension is easily extendable. As an illustration, we show the impact of such improvements when applied to the real exchange rate convergence within industrialized countries. The DF-GLS-SUR test provides strong evidence of PPP where the other tests considered failed.

References

Banerjee, A., 1999, Panel Data Unit Root and Cointegration: An Overview, Oxford Bulletin of Economics and Statistics, 61, 607-629.

Bowman, D., 1999, Efficient Tests for Autoregressive Unit Roots in Panel Data, International Finance Discussion Papers, number 646.

Chang, Y., 2002, Bootstrap Unit Root in Panels with Cross-Sectional Dependency, Journal of Econometrics, forthcoming.

Elliott, G., T. Rothenberg, and Stock, J.H., 1996, Efficient Tests for an Autoregressive Unit Root, Econometrica, 64, 813-836.

Hall, A., 1994, Testing for a Unit Root in Time Series with Pretest Data-Based Model Selection, Journal of Business and Economic Statistics, 12, 461-470.

Hansen, B. E., 1995, Rethinking the Univariate Approach to Unit Root Testing: Using Covariates to Increase Power, Econometric Theory, 11, 1148-1172.

Im, K.S., M.H. Pesaran, and Y. Shin, 2003, Testing for Unit Roots in Heterogeneous Panels, Journal of Econometrics, 115: 53-74.

Levin, A., C.F. Lin, and C.J. Chu, 2002, Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties, Journal of Econometrics, 108, 1-24.

Lopez, C., Murray C.J., and D.H. Papell, 2004, State of the Art Unit Root Tests and Purchasing Power Parity, Journal of Money Credit and Banking, forthcoming.

Maddala, G.S. and S. WU, 1999, A comparative Study of Unit Root Tests with Panel Data and a New Simple Test, Oxford Bulletin of Economics and Statistics, 61: 631-652.

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

Ng. S., and P. Perron, 2001, Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power, Econometrica, 69, 1519-1554.

O'Connell P., 1998, The Overvaluation of Purchasing Power Parity, Journal of International Economics, 44, 1-19.

Papell, D.H., 1997, Searching for Stationarity: Purchasing Power Parity under the Current Float, Journal of International Economics, 43, 313-332.

Papell, D.H., and H. Theodoridis, 2001, The Choice of Numeraire Currency in Panel Tests of Purchasing Power Parity, Journal of Money, Credit and Banking, 33, 790-803.

Table 1: Size-Adjusted Power for the ADF-SUR and the DF-GLS-SUR Tests, z=(1)

N	T	ρ = 0.99		0.97		0.95		0.90	
		ADF-SUR	DF-GLS-SUR	ADF-SUR	DF-GLS-SUR	ADF-SUR	DF-GLS-SUR	ADF-SUR	DF-GLS-SUR
5	25	0.0588	0.1246	0.0726	0.2836	0.0870	0.4434	0.1660	0.7968
	50	0.0646	0.1782	0.1002	0.5596	0.1668	0.8286	0.5156	0.9968
	75	0.0848	0.2814	0.1810	0.8032	0.3562	0.9752	0.8812	1.0000
	100	0.0868	0.3766	0.2310	0.9422	0.5342	0.9980	0.9894	1.0000
	125	0.1082	0.4718	0.3426	0.9700	0.7400	1.0000	0.9994	1.0000
10	25	0.0622	0.1686	0.0750	0.4646	0.1022	0.7112	0.2132	0.9708
	50	0.0720	0.3460	0.1410	0.8894	0.2790	0.9944	0.7956	1.0000
	75	0.1028	0.5052	0.2688	0.9860	0.5876	1.0000	0.9942	1.0000
	100	0.1160	0.6710	0.4136	0.9996	0.8474	1.0000	1.0000	1.0000
	125	0.1440	0.7832	0.5970	1.0000	0.9664	1.0000	1.0000	1.0000
15	25	0.0542	0.2198	0.0674	0.5932	0.0778	0.8280	0.1454	0.9954
	50	0.0794	0.4666	0.1718	0.9730	0.3614	0.9998	0.9148	1.0000
	75	0.1154	0.6906	0.3624	0.9996	0.7494	1.0000	0.9994	1.0000
	100	0.1456	0.8178	0.5812	1.0000	0.9528	1.0000	1.0000	1.0000
	125	0.1556	0.9350	0.9904	1.0000	0.9938	1.0000	1.0000	1.0000
20	25	-	-	-	-	-	-	-	-
	50	0.0896	0.5578	0.1860	0.9912	0.4044	1.0000	0.9538	1.0000
	75	0.1284	0.7998	0.4342	1.0000	0.8456	1.0000	1.0000	1.0000
	100	0.1506	0.9350	0.6366	1.0000	0.9808	1.0000	1.0000	1.0000
	125	0.2092	0.9772	0.7294	1.0000	1.0000	1.0000	1.0000	1.0000

Table 2: Size-Adjusted Power for the ADF-SUR and the DF-GLS-SUR Tests, z=(1,t)

N	T	ρ = 0.99		0.97		0.95		0.90	
		ADF-SUR	DF-GLS-SUR	ADF-SUR	DF-GLS-SUR	ADF-SUR	DF-GLS-SUR	ADF-SUR	DF-GLS-SUR
5	35	0.0504	0.0520	0.0530	0.0638	0.0652	0.0804	0.1050	0.1730
	50	0.0510	0.0544	0.0616	0.0690	0.083	0.1012	0.2850	0.2906
	75	0.0514	0.0576	0.0830	0.0984	0.1412	0.1880	0.4960	0.6540
	100	0.0594	0.0638	0.0958	0.1418	0.2098	0.3330	0.7544	0.9144
	125	0.0596	0.0640	0.1400	0.1860	0.3412	0.4740	0.9486	0.9866
10	35	0.0542	0.0522	0.0550	0.0576	0.0762	0.0894	0.1518	0.2502
	50	0.0550	0.0524	0.0774	0.0874	0.1080	0.1452	0.3400	0.5238
	75	0.0556	0.0598	0.0964	0.1176	0.2050	0.2884	0.7604	0.9142
	100	0.0598	0.0644	0.1456	0.2002	0.3670	0.5412	0.9724	0.9966
	125	0.0668	0.0724	0.2058	0.3076	0.5768	0.7840	1.0000	1.0000
15	35	0.0540	0.0548	0.0636	0.0732	0.0808	0.1150	0.1762	0.3654
	50	0.0538	0.0568	0.0748	0.0896	0.1204	0.1796	0.4466	0.6810
	75	0.0556	0.0588	0.1096	0.1530	0.2542	0.4002	0.8896	0.9838
	100	0.0588	0.0678	0.1656	0.2744	0.4608	0.7232	0.9966	1.0000
	125	0.0704	0.0844	0.2582	0.4304	0.7178	0.9100	1.0000	1.0000
20	35	-	-	-	-	-	-	-	-
	50	0.0508	0.0550	0.0688	0.0970	0.0084	0.2078	0.4596	0.7866
	75	0.0538	0.0586	0.1110	0.1778	0.2736	0.4850	0.9396	0.9960
	100	0.0648	0.0740	0.2008	0.3454	0.5688	0.8350	0.9990	1.0000
	125	0.0750	0.0920	0.3216	0.5276	0.8248	0.9738	1.0000	1.0000

Table 3: Empirical Size of the DF-GLS-SUR test, z=(1), AR(1) processes

N	T	$\omega =$ 0.25	0.5	0.75
5	25	0.0524	0.0550	0.0644
	50	0.0452	0.0464	0.0506
	75	0.0514	0.0528	0.0546
	100	0.0524	0.0528	0.0552
	125	0.0554	0.0568	0.0570
	250	0.0558	0.0548	0.0556
10	25	0.0484	0.0642	0.1050
	50	0.0592	0.0684	0.0950
	75	0.0560	0.0618	0.0736
	100	0.0526	0.0550	0.0644
	125	0.0510	0.0548	0.0610
	250	0.0506	0.0502	0.0532
15	25	0.0636	0.1030	0.1916
	50	0.0610	0.0844	0.1426
	75	0.0576	0.0698	0.1142
	100	0.0536	0.0596	0.0890
	125	0.0610	0.0688	0.0900
	250	0.0436	0.0464	0.0566
20	25	0.0690	0.1398	0.2634
	50	0.0658	0.1006	0.2120
	75	0.0584	0.0848	0.1644
	100	0.0646	0.0796	0.1426
	125	0.0534	0.0668	0.1174
	250	0.0560	0.0602	0.0778

Table 4: Empirical Size of the DF-GLS-SUR test, z=(1,t), AR(1) processes

N	T	$\omega =$ 0.25	0.5	0.75
5	35	0.0598	0.0634	0.0670
	50	0.0582	0.0610	0.0618
	75	0.0740	0.0494	0.0496
	100	0.0514	0.0514	0.0526
	125	0.0534	0.0534	0.0550
	250	0.0500	0.0500	0.0504
10	35	0.0622	0.0820	0.1310
	50	0.0570	0.0652	0.0992
	75	0.0452	0.0504	0.0684
	100	0.0536	0.0570	0.0726
	125	0.0476	0.0516	0.0606
	250	0.0528	0.0544	0.0574
15	35	0.0696	0.1232	0.2186
	50	0.0690	0.1026	0.1774
	75	0.0484	0.0612	0.1100
	100	0.0558	0.0662	0.0920
	125	0.0514	0.0576	0.0788
	250	0.0552	0.0588	0.0666
20	35	0.0838	0.1784	0.3192
	50	0.0736	0.1354	0.2466
	75	0.0496	0.0850	0.1688
	100	0.0522	0.0806	0.1418
	125	0.0532	0.0694	0.1118
	250	0.0554	0.0600	0.0792

Table 5: Empirical Size of the DF-GLS-SUR test, AR(p) processes

T	N	*Panel*	Size
25	20	*All*	0.1930
	12	*13*	0.0530
	11	*EC*	0.1196
	11	*E11*	0.0828
	10	*G10*	0.0694
	10	*E10*	0.0832
	7	*EMS*	0.0728
	6	*G6*	0.0570
50	20	*All*	0.0912
	12	*13*	0.0496
	11	*EC*	0.0878
	11	*E11*	0.0704
	10	*G10*	0.0674
	10	*E10*	0.0674
	7	*EMS*	0.0518
	6	*G6*	0.0486
100	20	*All*	0.0632
	12	*13*	0.0438
	11	*EC*	0.0530
	11	*E11*	0.0556
	10	*G10*	0.0404
	10	*E10*	0.0602
	7	*EMS*	0.0538
	6	*G6*	0.0784

Table 6: Univariate Unit Root Tests

	ADF			DF-GLS		
	α	t_α	k^r	α	t_α	k^{MAIC}
Australia	-0.0433	-1.3703	1	-0.0344	-1.4520	1
Austria	-0.0610	-2.0064	1	-0.0491	-1.5233	5
Belgium	-0.0590	-2.1222	3	-0.0477	-1.8053 *	1
Canada	-0.0225	-1.2412	4	-0.0052	-0.3448	4
Denmark	-0.0670	-2.2189	3	-0.0427	-1.6792 *	1
Finland	-0.1146	-2.8181 *	7	-0.0375	-1.5823	1
France	-0.0846	-2.3967	7	-0.0662	-2.2368 **	1
Germany	-0.0751	-2.2518	7	-0.0598	-1.9829 **	1
Greece	-0.0714	-2.3391	4	-0.0550	-1.8926 *	5
Ireland	-0.0974	-2.4920	1	-0.0658	-2.0391 **	1
Italy	-0.0865	-2.4702	3	-0.0863	-2.4847 **	3
Japan	-0.0409	-1.7202	5	-0.0147	-0.9587	1
Netherlands	-0.0824	-2.3577	7	-0.0528	-1.8938 *	1
Norway	-0.1073	-2.6052 *	7	-0.0559	-1.8912 *	1
New Zealand	-0.1111	-2.5669	9	-0.1223	-3.3023***	6
Portugal	-0.0556	-2.0349	4	-0.0459	-1.8552 *	4
Spain	-0.0560	-2.0832	2	-0.0316	-1.5516	2
Sweden	-0.0813	-2.5893 *	7	-0.0491	-1.7399 *	1
Switzerland	-0.0837	-2.5483	3	-0.0282	-1.3312	1
U.K	-0.10948	-2.5152	6	-0.0477	-1.4565	0

*,**,** indicate respectively evidence of PPP at 10%, 5% and 1%.

Table 7: Multivariate Unit Root Tests

	ADF - SUR		DF - GLS - SUR	
	α	t_α	α	t_α
All20	0.910	-8.400 **	0.977	-4.884***
EC	0.929	-6.777***	0.958	-5.027***
EMS	0.931	-5.249 **	0.955	-4.105***
G6	0.944	-3.760	0.975	-2.835***
G10	0.944	-5.598 **	0.974	-3.675 ***
Euro10	0.945	-5.737 **	0.985	-3.078 **
Euro11	0.945	-6.043 *	0.978	-3.339***
13	0.936	-6.790***	0.969	-4.431***

*,**,** indicate respectively evidence of PPP at 10%, 5% and 1%.

Figure 1: Robustness Analysis

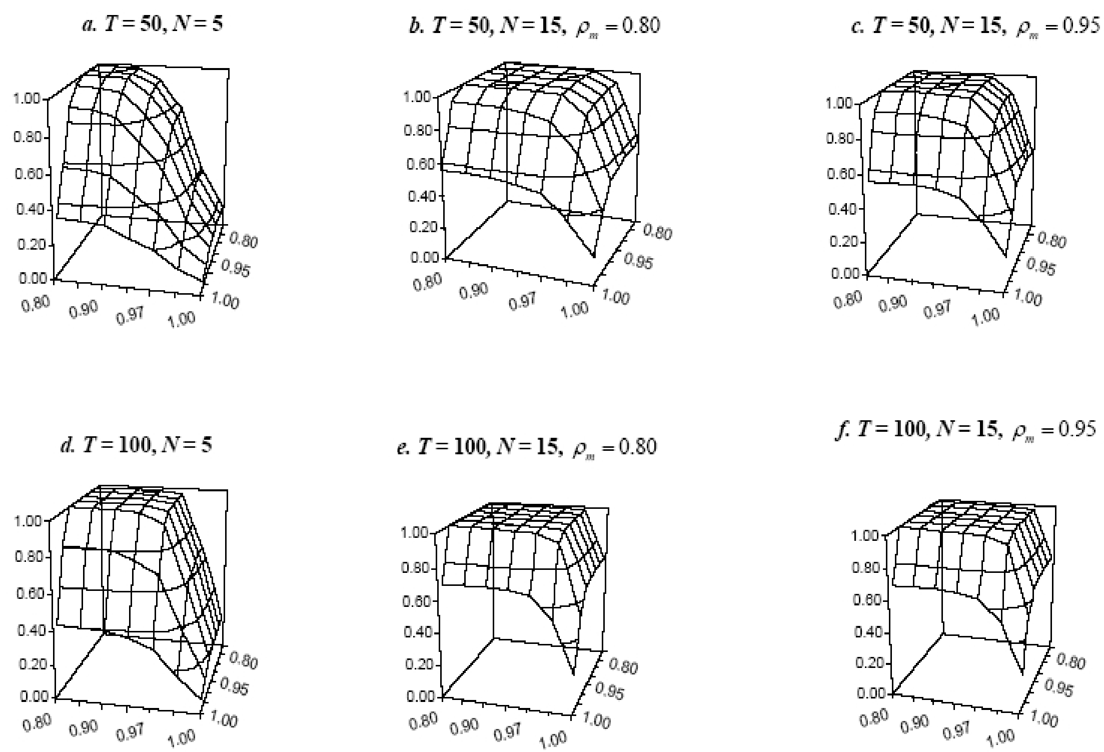

Figure 2: The DF-GLS-SUR Test (—) Versus The IPS Test (−−)