Panel Unit Root Tests with GLS-Detrending With an Application to Purchasing Power Parity

July 2003

Claude Lopez
University of Cincinnati

We combined recent developments in univariate and multivariate unit root testing in order to construct a more-powerful panel unit root test. We extended the GLS-detrending procedure of Elliott, Rothenberg, and Stock (1996) to a panel Augmented Dickey-Fuller test. The finite sample power properties of the new test demonstrate a very large gain when compared to existing tests, especially for small panels. We then investigated the topic of Purchasing Power Parity for the post Bretton-Woods period via this new test. The results show strong rejections of the unit root test hypothesis.

Correspondence to: Claude Lopez, Department of Economics, University of Cincinnati, E-mail (temporary): Claude.Lopez@mail.uh.edu

I. Introduction

Even though most international economic theoretical models imply or require the stationarity of macroeconomics time series, such as real exchange rates or output, the implementation of this concept remains a sensitive matter due to the inadequacy of most of the tools available. Commonly, the stationarity is defined as a long-run relationship toward which the series are supposed to converge. A suitable method to address this issue is to investigate whether or not the deviations from the level of equilibrium are permanent, via unit root tests.

The augmented Dickey-Fuller (ADF) unit root test is the most well-known of these tests. However, this univariate method demonstrates poor power properties, especially for the length of data sets available, which are widely documented and have led to burgeoning literature attempting to overcome these disadvantages. These developments have occurred at both multivariate and univariate levels.

At the multivariate level, several panel unit root tests have been developed, which augment time-series information with cross-sectional variability. Among these are the tests of Levin, Lin, and Chu (LLC) (2002), Im, Pesaran, and Shin (1997), and Maddala and Wu (1996).[1] These authors have demonstrated that, for as few as 5 time-series, their panel unit root tests have significantly higher power than the univariate unit root tests available.

The LLC test and, more specifically, the LLC hypothesis are the most widely used. Indeed, several works propose enhanced versions of this test, producing more 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] Finally, Papell and Theodoridis (2001) acknowledge these adjustments by considering a panel version of the ADF test, using the LLC hypothesis. They allow for heterogeneous serial *and* contemporaneous correlation.

There have also been important developments at the univariate level. Elliott, Rothenberg, and Stock (1996) developed a GLS-detrended/demeaned version of the ADF

[1] Banerjee (1999) provides an exhaustive survey of the major tests.
[2] Papell (1997) shows a strong relation between the size of the panel and the rate of rejection of the unit root.

test. Running the ADF test on the GLS-transformed data leads to one of the most powerful univariate tests, which is called the DF-GLS test.[4]

Both the panel approach and the DF-GLS test deliver a substantial gain in power over the univariate ADF unit root test, as well as more rejections of the unit root hypothesis at the empirical level. Using a panel version of the ADF test, Papell and Theodoridis (2001) offer some evidences of PPP, depending on the numeraire and the width of the panel considered. For the same period, Cheung and Lai (2000) are able to reach more evidences of PPP through the DF-GLS test, rather than the ADF test, for the industrialized countries. Overall, while these results are encouraging, they remain too weak to allow any clear conclusion.

The aim of this work is to provide a better technique to test the potential existence of unit roots and to measure if this improvement has any impact at the empirical level. Seeking a significant increase in power over existing tests, we combined the GLS-detrending of Elliott, Rothenberg, and Stock (1996) with the panel ADF test, using Levin, Lin, and Chu's (2002) hypotheses. We analyzed the behavior of our panel unit root test for various sample sizes, panel widths, and degrees of persistence with a Monte Carlo experiment. The main result is that our panel unit root test displays significantly better finite sample power than existing univariate and panel unit root tests.

We then applied this new panel unit root test to a data set where neither the panel ADF test nor the DF-GLS test are able to reject strongly the unit root: the post Bretton-Woods real-exchange rates of the industrialized countries. The principal outcome is robust, overall support for the PPP hypothesis, independent of the width and the length of the panel considered.

The next section provides a non-exhaustive review of the literature on unit root tests, while Section III develops the new panel unit root test and tabulates the finite sample critical values. Section IV conducts a detailed power experiment, where it is shown that the new test provides a significant increase in power over the panel ADF test. Section V presents an empirical application to PPP, where it is demonstrated that the new test leads to strong

---

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

[4] Still at the univariate level, Hansen (1995) proposes a more-powerful alternative to the DF-GLS test by including covariates to the test.

rejections where the panel ADF test has failed. Finally, Section VI summarizes our findings and offers concluding remarks.

II. Existing Unit Root Tests: A Review of the Literature

The standard ADF unit root test 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 the unit root null hypothesis is that $\alpha = 1$ and the alternative of stationarity is $\alpha < 1$. d_t is a set of deterministic regressors, and k lagged first difference terms are included to account for serial correlation. This test is well-known for its poor power, and subsequent literature suggests several solutions. In the next two subsections, we describe some recent developments at both multivariate and univariate levels.

IIa. 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 unit roots 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 a homogenous alternative in which every series in the panel is stationary with the same speed of reversion. Im, Pesaran and Shin (1997) and Maddala and Wu (1996) test the unit root null hypothesis against the alternative in which *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 the value α is constrained to be the same for each series in the panel. The null hypothesis is that $\alpha = 1$, and the alternative is that $\alpha < 1$. For each series j, $j = 1, ..., N$,. $d_{jt} = \beta_j ' z_t$ is a set of deterministic regressors, which allows for heterogeneous intercepts and time trends, and k_j lagged first difference terms 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 (1996) 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, which we refer to as 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; that is, we shall use the ADF-SUR test.

Performing the ADF-SUR test is a two-step procedure. First, for each series *j*, $j = 1, ..., N$, we must select the number of lagged first difference terms, k_j, to account for serial correlation. We use the general-to-specific lag selection procedure of Hall (1994) and Ng and Perron (1995). Then, having selected k_j, we estimate the system of equations using SUR, and constrain the value of α to be equal across equations.

IIb. 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:

- The demeaned case, $z_t = (1)$:

$$y_t^m = \alpha y_{t-1}^m + \sum_{i=1}^{k} \psi_i \Delta y_{t-i}^m + u_t . \tag{3}$$

- The detrended case, $z_t = (1,t)$:

[5] However, O'Connell (1998) imposes homogeneous serial correlation properties across the series, which results in under rejection of the null.

$$y_t^t = a y_{t-1}^t + \sum_{i=1}^{k} y_i \Delta y_{t-i}^t + u_t \qquad (4)$$

where y_t^m (y_t^t) is the GLS-demeaned (GLS-detrended) series.

Equations (3) and (4) can be rewritten as:

$$y_t^{GLS} = a y_{t-1}^{GLS} + \sum_{i=1}^{k} y_i \Delta y_{t-i}^{GLS} + u_t \ , with\, GLS = (m,t) \qquad (5)$$

with $y_t^{GLS} = y_t - \tilde{b} z_t$, $GLS = (m,t)$, $\tilde{b}$ is the least squares estimate of the regression of $\tilde{z}_t$ on $\tilde{y}_t$, *i.e.* $\tilde{b} = \left(\sum \tilde{z}_t^2\right)^{-1} \sum \tilde{z}_t \tilde{y}_t$. $\tilde{y}_t$ and $\tilde{z}_t$ are the quasi-differences of y_t and z_t respectively, i.e. $\tilde{y}_t = (y_1,(y_2 - a y_1),...,(y_T - a y_{T-1}))'$, and $\tilde{z}_t = (z_1,(z_2 - a z_1),...,(z_t - a z_{t-1}))'$. $a = 1 + \bar{c}/T$ represents the local alternative, with $\bar{c} = -7$ when $z_t = (1)$ and $\bar{c} = -13.5$ when $z_t = (1,t)$.[6]

The issue of the lag selection 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), which 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.

III. Panel Unit Root Tests with GLS-Detrending: The DF-GLS-SUR Test

While both the ADF-SUR and the DF-GLS tests demonstrate higher power than the standard ADF test, they display a limited ability to reject the unit root null for macroeconomic 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, which we refer to as the DF-GLS-SUR test, runs the following panel version of equation (5):

$$y_{jt}^{GLS} = a y_{j,t-1}^{GLS} + \sum_{i=1}^{k_j^{MAIC}} y_{ji} \Delta y_{j,t-i}^{GLS} + u_{jt}, \ with\, GLS = (m,t) \qquad (6)$$

[6] $\bar{c}$ = -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 constant (the case with constant and 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.

For each series j, $j=1,\ldots,N$, k_j represents the number of lagged first difference terms, allowing for serial correlation. Similarly to the ADF-SUR test, the DF-GLS-SUR test is performed through a two-step procedure. For each series j in the panel, we first GLS-transform the data and select k_j, using MAIC. Then, having selected k_j, we estimate the system of equations via SUR, and constrain the value of α to be equal across equations. We expect that combination of those two techniques will lead to a summation of their respective power properties.

The ADF-SUR test represents the benchmark against which we compare the DF-GLS-SUR test. Even though Papell and Theodoridis (2001) use the ADF-SUR test, they do not report generic critical values. Therefore, we need to generate the finite sample critical values for both the ADF-SUR and the DF-GLS-SUR tests. The Monte Carlo experiment considers panels with a length of T = 25, 50, 75, 100, and 125, and with a width of N = 5, 10, 15, and 20.[8] The resulting panels (N, T) represent the most commonly examined sizes in macroeconomics. The data are generated under the null hypothesis as random walks without drift:

$$y_{jt} = y_{j,t-1} + u_{jt}, \qquad (7)$$

where $u_{jt} \sim iidN(0,1)$ and the error terms are contemporaneously uncorrelated, $E(u_{it}u_{jt}) = 0$ for $i \neq j$.[9]

For each panel unit root test, we generate four sets of critical values, two sets for each model (regression with constant only and with constant and trend). First, we assume that the true lag length is known; that is, we fix k = 0. We are also interested in the effects of lag selection on the finite sample distribution of the unit root test statistics. Accordingly, we generate critical values where at each iteration, we select the lag length by GS for the ADF-SUR test and by MAIC for the DF-GLS-SUR test.

In accordance with Hall's (1994) findings, our results demonstrate that the inclusion of the serial correlation, selected via the GS procedure, induces a strong increase in absolute value of the critical values. Furthermore, the percentage change in critical values is more

[8] We consider T=(35, 50, 75, 100, 125) for the case with heterogeneous constants and trends.
[9] In the subsequent empirical exercise, we shall explicitly allow for contemporaneously correlated errors.

severe for the ADF-SUR test than for the DF-GLS-SUR test, i.e. when the lag is selected via GS instead of via MAIC. With the true value of k being 0 and the MAIC procedure providing the best estimation of the lag length, the critical values for both cases, $k = 0$ and $k = k^{MAIC}$, are relatively close.

The 1%, 5%, and 10% critical values are reported in Tables 1a, and 1c for the ADF-SUR test and *1b* and *1d* for the DF-GLS-SUR test.

IV. Finite Sample Performance: Power Analysis

Because it is a combination of the ADF-SUR and the DF-GLS tests, we expect the DF-GLS-SUR test to be more powerful than each one of them. To be able to compare the performance of these three tests, we need to compute the power of the ADF-SUR and the DF-GLS-SUR tests.[10] Considering the same panels (N, T) for the critical values, the power is computed via a Monte Carlo experiment with the data generated under the alternative, that is:

$$y_{jt} = \rho y_{j,t-1} + u_{jt} \qquad (8)$$

where $u_{jt} \sim iidN(0,1)$, the error terms are contemporaneously uncorrelated, $E(u_{it}u_{jt}) = 0$ for $i \neq j$, and $\rho < 1$. We consider the following alternatives $\rho = (0.99, 0.97, 0.95, 0.90, 0.85, 0.80)$, with the nominal size fixed at 5%. Tables 2a, 2b, 2c and 2d display the level of power for the ADF-SUR and the DF-GLS-SUR tests, for the ($k = 0$) and the ($k = (k^{GS}, k^{MAIC})$) cases using regression with constants only (*2a* and *2b*) or with constants and trends (*2c* and *2d*).

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.

IVa. Demeaned Case: Tables 2a and 2*b*

A first 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 improvements in power: 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 0.98, with $N = 5$.

[10] Elliott, Rothenberg, and Stock (1996) provide the power analysis for the DF-GLS test for both demeaned and detrended cases.

Commonly, the lag selection induces a uniform loss in power, compared to the case where the lag length is known and is equal to 0. The simulations show this expected outcome, with Table 2b having a lower power than Table 2a. Except for this difference, however, Tables 2a and 2b present a similar pattern: the DF-GLS-SUR test demonstrates an overall higher power than the ADF-SUR test. Therefore, the following performance analysis does not dissociate these two cases, unless it is clearly specified.

A sole increase in T leads to a consistent increase in power for both tests, with significantly stronger improvements for the DF-GLS-SUR test than for the ADF-SUR test. For example, considering the case with no lags, a highly persistent system of series, $\boldsymbol{r} = 0.99$, 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 $\boldsymbol{r} = (0.97, 0.95)$, the DF-GLS-SUR test continues to present a stronger response to an increase in the number of observations than the ADF-SUR test.

Considering increases only for the number of series, N, we also observe consistent improvements in power, with a stronger impact on the DF-GLS-SUR test than on the ADF-SUR test. For example, with lag selection, $\boldsymbol{r} = 0.99$, and $T = 25$, an increase in the number of series, N evolving from 5 to 10, leads to an increase in power for the DF-GLS-SUR test five times stronger than for the ADF-SUR test. A change in N induces a stronger response from the DF-GLS-SUR test than the ADF-SUR test. Furthermore, by comparing the increase in power due to a change in T with the increase due to a change in N, the tables for both tests show that an increase in width has a stronger impact on power than an increase in length.[11] Considering the case with no lag selection and $\boldsymbol{r} = 0.95$, 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 level of power close to 1, while the addition of only 5 series displays the same result.

[11] LLC, Maddala, and Wu (1996) and Im, Pesaran, and Shin (1997) demonstrate the same 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. In the presence of lag selection, and with r = 0.99, this combination presents a power of 0.44 for the DF-GLS-SUR test and of 0.10 for the ADF-SUR test. If r = 0.97, the DF-GLS-SUR test reaches a power of 0.98, while the ADF-GLS test offers only a power of 0.22. The DF-GLS-SUR test demonstrates an impressive higher power than the ADF-SUR test for panels that include highly persistent series and a width close to the length.

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. In the case with lag selection, and (N, T) = (10, 75), for example, the DF-GLS-SUR test presents a power of 0.41 for r = 0.99, of 0.98 for r = 0.97, and of 1.00 for r = 0.95. More generally, the DF-GLS-SUR test has a power of 1.00 or close to 1.00 whenever $r < 0.90$.

IVb. Detrended Case: Tables 2c and 2d

Commonly, the addition of a trend to the regressions leads to a uniform loss in power for both tests: with r = 0.97 and (N, T) = (15, 100), the DF-GLS-SUR^m test produces a power of 1.00, while the DF-GLS-SUR^t test reaches only a power of 0.30. However, combining the time-series information with cross-sectional information still provides significant improvements in the test performance. For r = 0.95 and T = 100, the standard DF-GLS^t test reaches a power level of 0.10, while the DF-GLS-SUR^t test, accounting for four more series, is able to achieve a power level of 0.35.

The GLS-transformation shows a similar impact on the size-adjusted power than in the previous section. If r = 0.95 and (N, T) = (5, 125), the DF-GLS-SUR test has a power of 0.53, while the ADF-SUR test offers a power level of 0.24.

Furthermore, the amplitude of these enhancements varies following changes in T, length of the data, in N, width of the panel or in r, persistence of the processes, as well as the test considered. For r = 0.97 and N = 20, a raise in T from 50 to 100 observations induces an increase in power of 0.23 for the DF-GLS-SUR test and of 0.10 for the ADF-SUR test.

Likewise, if N varies from 15 to 20 series, when $\rho = 0.97$ and $T = 100$, the power augments by 0.32 for the DF-GLS-SUR test and by 0.03 for the ADF-SUR test. Finally, a decrease in the persistence, from $\rho = 0.97$ to $\rho = 0.95$, generates strong improvements in performance for both tests: for $(N, T) = (10, 125)$, the observed increase in power is of 0.45 for the DF-GLS-SUR test and of 0.25 for the ADF-SUR test.

To sum up, this analysis reveals two major outcomes. First, as expected, by incorporating cross-sectional variation, we are able to dramatically enhance the power of the univariate DF-GLS test. Secondly, the comparison of both tests performances demonstrates the strong improvements due to the GLS-transformation. Overall, the DF-GLS-SUR test has higher finite sample power than the ADF-SUR test, and for each increase in information (either N or T), the corresponding increase in power is higher for the DF-GLS-SUR test than for the ADF-SUR test. Furthermore, our new test presents some interesting features: its power is attractively high for small panels and for highly-persistent series.

IVc. Robustness Analysis

One obvious objection to this new test stands in the homogeneity imposed by the alternative hypothesis, $\rho < 1$. This restriction seems to undermine the enhanced power properties previously presented. As a result, this section focuses on the impact of such a constraint by measuring the performance of the DF-GLS-SUR test when applied to series with heterogeneous rates of convergence, ρ_j.

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:

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

where $u_{jt} \sim iidN(0,1)$, the error terms are contemporaneously uncorrelated, $E(u_{it}u_{jt}) = 0$ for $i \neq j$, and $\rho_j \leq 1$. Then, we estimate equation (3) with $k = 1$.[12]

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

[12] This allows us to compare the size-adjusted power with Bowman (1999) and IPS (1997).

- N = 5, T = (50, 100) and r_j such that r_i = (1.00, 0.99, 0.97, 0.95, 0.90, 0.85, 0.80) with i = 1, 2, 3 and r_l = (1.00, 0.99, 0.97, 0.95, 0.90, 0.85, 0.80) with l = 4, 5.
- N = 15, T = (50, 100) and r_j such that r_i = (1.00, 0.99, 0.97, 0.95, 0.90, 0.85, 0.80) with i = 1, 2, 3, 4, 5, r_l = (1.00, 0.99, 0.97, 0.95, 0.90, 0.85, 0.80) with l = 6, 7, 8, 9, 10 and r_m = (0.80, 0.95) with m = 11, 12, 13, 14, 15.

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

The 3D graphs demonstrate strong deteriorations in the DF-GLS-SUR test performance in presence of random walks among the series. For example, if *(N, T)* = (5, 50) and (r_i, r_l) = (1.00, 0.99), the power level reached is 0.08 instead of 0.18, when (r_i, r_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 (r_i, r_l, r_m) = (0.97, 0.99, 0.80), instead of 1.00 when (r_i, r_l, r_m) = (0.80, 0.80, 0.80). The converse is also verified: if (r_i, r_l, r_m) = (0.85, 0.90, 0.95), the power equals 1.00, while if (r_i, r_l, r_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 IPS test. Figure 2 graphs the power simulations such that the X-axis represents the number of stationary series among the panel, and the Y-axis is the power. The panels considered have a width of N = 5, 10, 15, 20 and a length fixed to T = 100. The rates of convergence for the stationary processes are r = (0.80, 0.90, 0.95).

Overall, the DF-GLS-SUR test demonstrates a higher power than the IPS test only when r =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.

The DF-GLS-SUR test alternative hypothesis has a limited impact on the test performance in presence of series converging at different rates, but rather a strong and negative effect when the panel combines stationary and non-stationary processes. However, the latter observation has no impact on the test reliability. Indeed, the DF-GLS-SUR test focuses on the stationarity of the *entire* panel: the presence of at least one unit root should lead to no rejection of the null hypothesis.[13]

To sum up, the DF-GLS-SUR test was designed to answer more accurately whether or not the panel converges. Its overall satisfying performance in the presence of homogeneous or heterogeneous rates of convergence in a stationary data set confirms its accuracy. Furthermore, the relatively low power achieved in the presence of random walks in the panel is not a major issue, because it is still significantly higher than the nominal size (5%).[14]

V. Empirical Evidence: Purchasing Power Parity

PPP is a fundamental assumption for most of the theoretical models in international macroeconomics. Since a direct consequence is the real exchange rates stationarity, unit root tests are commonly used to analyze this issue. However, the empirical evidence of PPP remains ambiguous because of the unit root tests power deficiency and the limited data available.

The theoretical innovations, some of them reviewed in the previous sections, create new incentives for empirical researchers investigating the topic of PPP. We observed three major movements in this growing literature. First, Frankel (1986), and Lothian and Taylor (1996) extended the span of the data, considering an overall long-run convergence through different nominal exchange rate regimes and are able to reject the unit root hypothesis. However, Engel (2000) finds some large size distortions for those long-run tests, which, he claims, cause spurious rejections. Secondly, some studies apply the ERS (1996) test, among

[13] Differently, the IPS test is supposed to reject the unit root hypothesis if at least one process in the panel is stationary.

[14] The only issue could be to over reject the null hypothesis, but by controlling for the size, i.e. the tendency to over-reject the null, we solve this issue as long as the power is significantly higher than the size.

them Cheung and Lai (2000), Taylor (2002) and Lopez, Murray, and Papell (LMP) (2002).[15] Yet, the strength of the results relies strongly on the lag selection and the span of the data.[16]

Finally, the third group comprises studies that apply the LLC (2002) framework, among them Frankel and Rose (1996), Jorion and Sweeney (1996), MacDonald (1996), Oh (1996), Wu (1996), and Papell (1997). Besides the increase in the information utilized, the strength of the empirical evidence depends on the panel size (rejections weaken whenever there are fewer than 10 countries), the data frequency (monthly data give stronger results than quarterly), and the period considered. Furthermore, the commonly assumed cross-sectional independency between the different real exchange rates within the panel brings the validity of some of the results into question. Relaxing this restriction and using large panels, Papell and Theodoridis (2001) show a strong mean reversion in real exchange rate behavior, as well as its dependence on the numeraire chosen and on the panel width.

Even if some of the studies cited reach conclusions in favor of convergence toward PPP, most of them imply unrealistic assumptions (cross-sectional independency). As an alternative, we ran the most powerful unit root test available, the DF-GLS-SUR test, allowing for serial and contemporaneous correlation, on post Bretton-Woods real exchange rates. More precisely, we focused on industrialized countries and considered only the Cassel version of PPP that is reversion to a constant mean.

We considered quarterly CPIs and nominal exchange rates in dollars, from 1973, first quarter, to 2001, fourth quarter (Source IFS, CD-ROM for 03/2002) for 21 industrialized countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Ireland, Italy, Japan, Netherlands, Norway, New Zealand, Portugal, Spain, Sweden, Switzerland, the U.K., and the U.S. For the European countries switching to the Euro, we collected the nominal exchange rate currency by U.S. Dollar from 1973 (1) to 1998 (4), then Euro by U.S. Dollar and used the official rate to convert into currency by U.S. Dollar. [17, 18] We then constructed the real exchange rate, q (in logarithm) following:

[15] Using over 100 years of data for 20 countries, the DF-GLS test and Lagrange Multiplier lag selection, Taylor (2002) obtains strong rejections of the unit root. Using the same data but the MAIC lag selection, Lopez, Murray, and Papell (2002) provide weaker results.
[16] Ng and Perron (2001) show the influence of the number of lags on the size distortion for the DF-GLS test.
[17] Euro countries are Austria, Belgium, Finland, France, Germany, Italy, Ireland, the Netherlands, Portugal, and Spain.

$$q = e + p^* - p$$

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

We first proceeded with univariate estimations of the real exchange rates through the ADF and the DF-GLS tests, using as lag selection the GS and the MAIC procedures respectively and for the entire period. The results are shown in Table 3. Few rejections of the unit root hypothesis are observed: the ADF test never rejects, while the DF-GLS test offers several rejections, varying from a 10% level for Denmark and Italy to a 5% level for Belgium, France, Germany, Greece, and the Netherlands.

Second, we estimated the real exchange rates at the multivariate level with the ADF-SUR and the DF-SUR-GLS tests. The ADF-SUR test is a version of the LLC test accounting for contemporaneous correlation. The inclusion of correlation among the errors invalidates the limit distribution of the LLC test and makes it unknown.[19] Maddala and Wu (1996) propose a bootstrapping alternative, and demonstrate that the LLC test offers good performance when this technique is used. Furthermore, Chang (2002) proves the asymptotic validity of this method. Therefore a Monte Carlo experiment is employed, allowing us to generate the critical values under the standard hypothesis, i.e. $H_0 : \alpha = 0$ versus $H_1 : \alpha < 0$.

In a previous section, we have described the estimation process, as well as the Monte Carlo experiment used to generate critical values. However, the data generating process used for the nonspecific analysis does not include the cross-sectional correlation. For the data-specific critical values, we need to estimate them by estimating the non-diagonal variance-covariance matrix of the innovations.

The real exchange rates can be defined such as:

$$q_{jt} = d_{jt} + \rho_j q_{j,t-1} + u_{jt} \qquad (9)$$

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 $i \neq j$.

We first ran ADF regressions for each series, using BIC lag selection in order to estimate the characteristics of each process. Those estimates are assumed to define the true data

---

[18]For Greece, the period is from 1973 Q1 to 1999 Q4.

[19] MW (1996), Banerjee (1999), Bowman (1999), and Chang (2002) point out this issue.

generating processes of x_{jt}. Then, we are able to deduce the u_{jt} and Σ, the variance-covariance matrix of the innovations, *i.e.* $(u_{1t}...u_{Nt}) \sim N(0_N, \Sigma)$. The unit root is imposed in the generated process by taking partial sums. Finally, we proceed with the rest of the Monte Carlo experiment: for each process, the estimation of equation (3) ((1)) selects k_j^{MAIC} (k_j^{GS}), then equation (6) ((2)) is estimated using SUR, with the pre-selected k_j^{MAIC} (k_j^{GS}). Repeating each procedure 5000 times creates a vector of statistics. Then, the critical values are calculated.

The data is grouped under several panels: 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 group these real exchange rates under 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).[20]

For each panel, we first estimated the period 1973(1) – 1988(1). We then add observations quarter-by-quarter, ending with the period 1973(1) – 2001(4). The critical values for the tests reflect the increasing span of the data. For both the ADF-SUR and the DF-GLS-SUR tests, the p-values from panels ending between 1988 and 2001 are graphed in Figure 3.

The panels considered vary in size, with a width including between 6 and 20 U.S.-real exchange rates and a length evolving from 60 to 116 observations. As shown in the performance analysis, the DF-GLS-SUR test demonstrates high power for these specific cases (a minimum power level of 70%), while the ADF-SUR test behaves poorly, at least for the small panels (a power level of 20%). For the panels studied, the bias of the DF-GLS-SUR test is negligible compared to the bias of the ADF-SUR test. Furthermore, the high power

[20] 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.

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 p-values are uniformly lower than the ADF-SUR-test one. Ultimately, the unit root null is rejected at, at least 5% for the 8 panels, while the ADF-SUR test presents results varying from no rejection for E10 to a rejection at 1% for EC. These contrasted results comply with the previous performance analysis: the lack of rejections observed with the ADF-SUR test derives from its power deficiency.

Besides the obvious improvements in term of rejections, the DF-GLS-SUR-test results offer other interesting aspects. With the same width (N = 11 and N = 10) but different real exchange rates, we obtain strong evidence of PPP since 1988 for EC and G10, while E11 and E10 observe a slower convergence toward a weaker rejection of the unit root null. Small panels, with 6 or 7 series, offer strong results. Empirical evidence coincides with the simulations: the strength of the results is independent from the size of the panel considered (N).[21]

Each graph represents the evolution of the results while the span of the data is increased. As expected, the ADF-SUR test is more sensitive to this supplement of information. However, an increase in the number of observations needs to be combined with a number of countries large enough (N = 20) to reach a power comparable to the one reached by the DF-GLS-SUR test. Then, both tests offer strong rejection of the unit root hypothesis.

As shown earlier, because the DF-GLS-SUR test results are more reliable than the ADF-SUR test one, we focused on them for a more detailed analysis. We first considered the period 1973(1) – 2001(4). Even though the results weaken starting at end of 1999 – beginning of 2000, evidence of PPP remains strong for all the panels. The unit root null is rejected at 1% for All20, EC, EMS, G10, and 13 and at 5% for G6, E10, and E11.

We, then, focused on the 1973(1) – 1998(2) period, narrowing the question from whether or not PPP holds for the post-Bretton-Woods period to whether or not it holds under the flexible nominal exchange rate regime. The rejections of the unit root null remain strong (at 1% for all the panels except E10, which rejects at 5%), demonstrating strong evidence of PPP for this period.

Finally, the DF-GLS-SUR test offers fewer variations in the strength of the results than the ADF-SUR test, which seems coherent with the absence of strong appreciation or depreciation of the dollar after 1988. The DF-GLS-SUR test provides the most evidence of PPP in the panel framework.

VI. Conclusion

The literature already provides several more-powerful alternatives to the ADF unit root test. However, these improved tests demonstrate limited ability to reject the unit root hypothesis when applied to macroeconomic time series. This paper attempts to produce a more efficient panel unit root test.

The DF-GLS-SUR test, an extension of Elliott, Rothenberg, and Stock's (1996) GLS-transformation to a version of the Levin, Lin and Chu's (2002) test, leads to a significantly more-powerful test. To sum up, for both demeaned and detrended cases, the DF-GLS-SUR test offers a uniformly-higher finite-sample power than the ADF-SUR test. Furthermore, the performance of the new test remains attractive when studying a data with heterogeneous rates of convergence across the series.

We applied the DF-GLS-SUR test to the topic of PPP for the post-1973 period. Considering panels with different lengths and widths, we can always reject the unit root hypothesis for the flexible nominal exchange rate regime (1973 - 1998) and for the entire period (1973 - 2001). The degree of rejection observed, consequent to the use of the DF-GLS-SUR test, provides strong evidence of the mean-reverting behavior of real-exchange rates. More specifically, the convergence toward PPP within the industrialized countries seems to be consistently strong for the period that started in 1973 and ended within the past 10 years. Only the relation between the US dollar and the Euro zone currencies presents a slower convergence, providing evidence that PPP only started in 1994. A weakening of the situation is observed starting at the end of 1999, which coincides with the world economic slow-down and the dollar depreciation.[22]

---

[21] Furthermore, Table 3 shows that the panel good results are not due to stationarity at the univariate level, as suggested by Taylor and Sarno (1998).
[22] Papell (2002) remarks that the PPP evidence strengthens when the dollars appreciates.

The DF-GLS-SUR test seems to offer interesting features, especially for highly-persistent processes. Both the simulations and the empirical application present this test as relevant, eventually, for any convergence issue in macroeconomics. Furthermore, if we focus on the empirical results, the evidence of PPP seems encouraging enough to justify an extension of the DF-GLS-SUR test framework to measure the persistence of the unit root in the real exchange rate.

References

Abuaf, N., and P. Jorion, 1990, "Purchasing Power Parity in the Long-Run," *Journal of Finance,* 45: 157-174.

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.

Cassel, Gustav, 1918, "Abnormal deviations in international exchanges," *The Economic Journal*, 28: 413-415.

Chang, Y., 2002 "Bootstrap Unit Root in Panels with Cross-Sectional Dependency," *mimeo*, Rice University.

Cheung, Y., and K.S. Lai, 2000, "On Cross-country Differences in the Persistence of Real Exchange Rates," *Journal of International Economics*, 50: 373-397.

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

Engel, C.M., 2000, "Long-Run PPP May Not Hold After All," *Journal of International Economics,* 51: 243-273.

Frankel, J., 1986, "International Capital Mobility and Crowding Out in the U.S. Economy: Imperfect Integration of Financial Markets or of Goods Markets?" in *How Open Is the U.S. Economy?*, edited by R. Hafer. Lexington Books.

Frankel, J., and A. Rose, 1996, "A Panel Project on Purchasing Power Parity: Mean Reversion Within and Between Countries," *Journal of International Economics*, 40: 209-224.

Jorion, P., and R. Sweeney, 1996, "Mean Reversion in Real Exchange Rates: Evidence and Implications for Forecasting," *Journal of International Money and Finance*, 15: 535-550.

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, 1997, "Testing for Unit Roots in Heterogeneous Panels," *mimeo,* University of Cambridge.

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., and D.H. Papell, 2002, "Convergence to Purchasing Power Parity at the Commencement of the Euro," *mimeo*, University of Houston.

Lopez, C., Murray C.J., and D.H. Papell, 2002, "State of the Art Unit Root Tests and the PPP Puzzle," *mimeo*, University of Houston.

Lothian, J., and M. Taylor, 1996, "Real Exchange Rate Behavior: The Recent Float from the Perspective of the Past Two Centuries," *Journal of Political Economy,* 104: 488-509.

Maddala, G.S. and S. WU, 1996, "A comparative Study of Unit Root Tests with Panel Data and a New Simple Test: Evidence from Simulations and Bootstrap," *mimeo*, Ohio State University.

Murray, C.J., and D.H. Papell, 2002, "The Purchasing Power Parity Persistence Paradigm," *Journal of International Economics,* 56: 1-19.

McDonald, R., 1996, "Panel Unit Root Tests and Real Exchange Rates," *Economics Letters*, 50: 7-11.

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.

Oh, K., 1996, "Purchasing Power Parity and Unit Root Tests Using Panel Data", *Journal of International Money and Finance*, 15: 405-418.

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

Papell, D.H., 2002, "The Panel Purchasing Power Parity Endgame," mimeo, University of Houton.

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.

Shiller, R., and P. Perron, 1995, "Testing for the Random Walk Hypothesis: Power versus Frequency of Observation," *Economics Letters*, 18: 382-386.

Taylor, A.M., 2002, "A Century of Purchasing Power Parity," *The Review of Economics and Statistics*, 84: 139-150.

Taylor, M. and L. Sarno, 1998, "The Behavior of Real Exchange Rates during the Post-Bretton Woods Period," *Journal of International Economics*, 46: 281-312.

Wu, Y., 1996, "Are Real Exchange Rates Nonstationary? Evidence from a Panel-Data Test," *Journal of Money, Credit and Banking*, 28: 54-63.

Table 1a: Finite Sample Critical Values for the ADF-SUR Test, $z_t = (1)$

$$\Delta y_{jt} = \beta_{0j} + \alpha y_{j,t-1} + \sum_{i=1}^{k_j^{GS}} \psi_{ji} \Delta y_{j,t-1} + u_{jt}$$

N	T	1%		5%		10%	
		$k=0$	$k=k^{GS}$	$k=0$	$k=k^{GS}$	$k=0$	$k=k^{GS}$
5	25	-5.6707	-7.2479	-4.8651	-6.1155	-4.4903	-5.5355
	50	-5.2441	-5.8627	-4.5867	-5.1077	-4.2463	-4.6495
	75	-5.0430	-5.4978	-4.4447	-4.7705	-4.1315	-4.4324
	100	-5.0184	-5.3612	-4.4298	-4.6821	-4.1170	-4.3136
	125	-4.9788	-5.2008	-4.3978	-4.5648	-4.1079	-4.2351
10	25	-7.6518	-9.3829	-6.6882	-8.1226	-6.2560	-7.4331
	50	-6.6553	-7.3540	-6.0092	-6.5780	-5.6690	-6.1478
	75	-6.4209	-6.8628	-5.7586	-6.1366	-5.4237	-5.7433
	100	-6.2951	-6.5348	-5.6789	-5.9484	-5.3702	-5.6094
	125	-6.2544	-6.5169	-5.6366	-5.8687	-5.3230	-5.5030
15	25	-10.4550	-11.5438	-9.1043	-10.1968	-8.3837	-9.5094
	50	-7.9136	-8.6581	-7.2513	-7.8649	-6.8559	-7.4345
	75	-7.5525	-8.0338	-6.8655	-7.2639	-6.5137	-6.8718
	100	-7.2861	-7.6344	-6.7148	-7.0169	-6.3795	-6.6551
	125	-7.2301	-7.5393	-6.7228	-6.8976	-6.3569	-6.5463
20	25	-	-	-	-	-	-
	50	-9.3641	-10.2436	-8.4795	-9.1318	-8.0607	-8.6413
	75	-8.5003	-8.9527	-7.8859	-8.2686	-7.5369	-7.8973
	100	-8.2865	-8.7111	-7.7323	-8.0245	-7.3488	-7.6499
	125	-8.1448	-8.4751	-7.5264	-7.7635	-7.1730	-7.3898

Table 1b: Finite Sample Critical Values for the DF-GLS-SUR Test, $z_t = (1)$

$$Dy^m_{jt} = ay^m_{j,t-1} + \sum_{i=1}^{k_j^{MAIC}} y_{ji} Dy^m_{j,t-i} + u_{jt}$$

N	T	1%		5%		10%	
		$k=0$	$k=k^{MAIC}$	$k=0$	$k=k^{MAIC}$	$k=0$	$k=k^{MAIC}$
5	25	-3.1576	-3.5736	-2.3489	-2.6501	-1.9262	-2.1687
	50	-2.9406	-2.9617	-2.2707	-2.3190	-1.8473	-1.9156
	75	-2.8451	-2.8570	-2.2057	-2.2081	-1.8601	-1.8896
	100	-2.8089	-2.8238	-2.1433	-2.1820	-1.7854	-1.8192
	125	-2.7992	-2.7987	-2.1297	-2.1258	-1.7666	-1.7813
10	25	-3.6376	-3.9895	-2.7942	-3.0074	-2.2337	-2.4728
	50	-3.1048	-3.2857	-2.4295	-2.5088	-2.0350	-2.1203
	75	-3.1182	-3.1464	-2.3658	-2.4636	-1.9687	-2.0469
	100	-2.9855	-2.9702	-2.3049	-2.3375	-1.8906	-1.9548
	125	-2.9169	-2.9290	-2.2864	-2.3249	-1.9417	-1.9620
15	25	-4.5752	-4.7473	-3.3616	-3.5269	-2.7309	-2.9017
	50	-3.4798	-3.5937	-2.6615	-2.7977	-2.2484	-2.3617
	75	-3.3142	-3.3346	-2.5281	-2.6392	-2.1707	-2.2413
	100	-3.1826	-3.2989	-2.4966	-2.5599	-2.0692	-2.1302
	125	-3.1182	-3.1900	-2.3693	-2.3957	-2.0246	-2.0595
20	25	-	-	-	-	-	-
	50	-3.8437	-3.9963	-2.9495	-3.1214	-2.4988	-2.6501
	75	-3.5204	-3.6599	-2.7059	-2.8150	-2.3362	-2.4182
	100	-3.3528	-3.4175	-2.5408	-2.6331	-2.1907	-2.2403
	125	-3.1817	-3.2041	-2.5374	-2.6082	-2.1950	-2.2489

Table 1c: Finite Sample Critical Values for the ADF-SUR Test, $z_t = (1,t)$

$$\Delta y_{jt} = \beta_{0j} + \beta_{1j} t + \alpha y_{j,t-1} + \sum_{i=1}^{k_j^{GS}} \psi_{ji} \Delta y_{j,t-1} + u_{jt}$$

N	T	1%		5%		10%	
		$k=0$	$k=k^{GS}$	$k=0$	$k=k^{GS}$	$k=0$	$k=k^{GS}$
5	35	-6.9862	-9.9797	-6.3261	-8.6032	-5.9711	-7.9052
	50	-6.7121	-7.8731	-6.1253	-7.1001	-5.7737	-6.6813
	75	-6.4433	-7.2291	-5.9397	-6.5950	-5.6530	-6.2486
	100	-6.4847	-7.0777	-5.9368	-6.4079	-5.6393	-6.0715
	125	-6.3935	-6.9050	-5.8432	-6.2775	-5.5704	-5.9402
10	35	-9.3890	-12.9343	-8.6539	-11.4676	-8.3111	-10.7788
	50	-8.8541	-10.4200	-8.2646	-9.5357	-7.9407	-9.0788
	75	-8.5459	-9.4320	-7.9619	-8.7541	-7.6502	-8.4123
	100	-8.4465	-9.1381	-7.8478	-8.4828	-7.5839	-8.1482
	125	-8.2961	-8.9418	-7.7640	-8.2326	-7.4833	-7.9433
15	35	-11.6544	-16.3897	-10.7663	-14.5835	-10.4220	-13.7604
	50	-10.7395	-12.3456	-10.1020	-11.5615	-9.7999	-11.0971
	75	-10.2063	-11.3239	-9.6669	-10.5951	-9.3861	-10.2325
	100	-10.0315	-10.8169	-9.4767	-10.1849	-9.1507	-9.8115
	125	-9.8206	-10.4819	-9.3496	-9.8900	-9.0727	-9.5505
20	35	-	-	-	-	-	-
	50	-12.5861	-14.5574	-11.9055	-13.6222	-11.5147	-13.1007
	75	-11.7610	-12.9922	-11.2252	-12.2593	-10.8897	-11.8815
	100	-11.4545	-12.2848	-10.8657	-11.6550	-10.5851	-11.2859
	125	-11.3289	-11.9535	-10.7024	-11.3388	-10.4153	-11.0085

Table 1d: Finite Sample Critical Values for the DF-GLS-SUR Test, $z_t = (1,t)$

$$\Delta y^t_{jt} = \alpha y^t_{j,t-1} + \sum_{i=1}^{k_j^{MAIC}} \psi_{ji} \Delta y^t_{j,t-i} + u_{jt}$$

N	T	1%		5%		10%	
		$k=0$	$k=k^{MAIC}$	$k=0$	$k=k^{MAIC}$	$k=0$	$k=k^{MAIC}$
5	35	-5.6655	-6.1837	-6.0890	-5.6667	-5.7906	-5.3943
	50	-5.4555	-5.7931	-5.3722	-4.8815	-5.0888	-4.6203
	75	-5.1615	-5.5635	-5.0796	-4.7250	-4.8120	-4.4606
	100	-5.0973	-5.4424	-4.9069	-4.6111	-4.6513	-4.3884
	125	-5.0596	-5.3895	-4.8384	-4.5767	-4.5512	-4.3201
10	35	-7.4675	-8.1771	-8.3094	-7.7066	-8.0535	-7.4251
	50	-7.0952	-7.6604	-7.1533	-6.5814	-6.9261	-6.3123
	75	-6.7551	-7.2354	-6.7399	-6.2724	-6.4708	-6.0011
	100	-6.5987	-6.9806	-6.4973	-6.1126	-6.2131	-5.8673
	125	-6.4887	-6.8641	-6.3360	-5.9645	-6.0661	-5.7220
15	35	-9.2049	-9.8818	-9.2049	-9.3956	-8.3701	-9.1696
	50	-8.4095	-9.1628	-8.6821	-7.9467	-8.4536	-7.6964
	75	-7.9856	-8.6046	-8.0940	-7.4962	-7.8286	-7.2521
	100	-7.7685	-8.1982	-7.7327	-7.2597	-7.4862	-7.0184
	125	-7.6625	-8.0551	-7.5585	-7.1493	-7.2943	-6.8944
20	35	-	-	-	-	-	-
	50	-10.5111	-9.7688	-10.0582	-9.2706	-9.8306	-9.0317
	75	-9.7579	-9.1550	-9.2963	-8.6639	-9.0418	-8.4335
	100	-9.3108	-8.8193	-8.8340	-8.3454	-8.5908	-8.1195
	125	-9.0688	-8.6311	-8.5981	-8.1462	-8.3455	-7.9277

Table 2a: Size-Adjusted Power for the ADF-SUR and the DF-GLS-SUR Tests, $z_t = (1)$**,** no lag selection, size 5%

N	T	*r* = 0.99		0.97		0.95		0.90		0.85		0.8	
		ADF-SUR	DF-GLS-SUR	ADF-SUR	DF-GLS-SUR	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	0.3036	0.9462	0.4998	0.9906
	50	0.0646	0.1782	0.1002	0.5596	0.1668	0.8286	0.5156	0.9968	0.8618	0.9998	0.9884	1.0000
	75	0.0848	0.2814	0.1810	0.8032	0.3562	0.9752	0.8812	1.0000	0.9984	1.0000	1.0000	1.0000
	100	0.0868	0.3766	0.2310	0.9422	0.5342	0.9980	0.9894	1.0000	1.0000	1.0000	1.0000	1.0000
	125	0.1082	0.4718	0.3426	0.9700	0.7400	1.0000	0.9994	1.0000	1.0000	1.0000	1.0000	1.0000
10	25	0.0622	0.1686	0.0750	0.4646	0.1022	0.7112	0.2132	0.9708	0.4422	0.9980	0.7128	1.0000
	50	0.0720	0.3460	0.1410	0.8894	0.2790	0.9944	0.7956	1.0000	0.9914	1.0000	1.0000	1.0000
	75	0.1028	0.5052	0.2688	0.9860	0.5876	1.0000	0.9942	1.0000	1.0000	1.0000	1.0000	1.0000
	100	0.1160	0.6710	0.4136	0.9996	0.8474	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	125	0.1440	0.7832	0.5970	1.0000	0.9664	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15	25	0.0542	0.2198	0.0674	0.5932	0.0778	0.8280	0.1454	0.9954	0.2866	0.9998	0.5420	1.0000
	50	0.0794	0.4666	0.1718	0.9730	0.3614	0.9998	0.9148	1.0000	0.9990	1.0000	1.0000	1.0000
	75	0.1154	0.6906	0.3624	0.9996	0.7494	1.0000	0.9994	1.0000	1.0000	1.0000	1.0000	1.0000
	100	0.1456	0.8178	0.5812	1.0000	0.9528	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	125	0.1556	0.9350	0.9904	1.0000	0.9938	1.0000	1.0000	1.0000	1.0000	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	1.0000	1.0000	1.0000	1.0000
	75	0.1284	0.7998	0.4342	1.0000	0.8456	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	100	0.1506	0.9350	0.6366	1.0000	0.9808	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	125	0.2092	0.9772	0.7294	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 2b: Size-Adjusted Power for the ADF-SUR and the DF-GLS-SUR Tests, $z_t = (1)$**,** with lag selection, size 5%

N	T	r = 0.99		0.97		0.95		0.90		0.85		0.8	
		ADF-SUR	DF-GLS-SUR	ADF-SUR	DF-GLS-SUR	ADF-SUR	DF-GLS-SUR	ADF-SUR	DF-GLS-SUR	ADF-SUR	DF-GLS-SUR	ADF-SUR	DF-GLS-SUR
5	25	0.0544	0.0710	0.0570	0.1682	0.0864	0.2330	0.1272	0.4470	0.1788	0.6152	0.2140	0.7126
	50	0.0622	0.1478	0.0846	0.4582	0.1514	0.6782	0.3442	0.9306	0.7360	0.9734	0.7362	0.9804
	75	0.0970	0.2520	0.1472	0.7412	0.2934	0.9206	0.6760	0.9932	0.9044	0.9992	0.9684	0.9994
	100	0.0856	0.3194	0.1938	0.8820	0.4288	0.9824	0.8842	0.9988	0.9832	1.0000	0.9976	1.0000
	125	0.1092	0.4350	0.3070	0.9278	0.6314	0.9970	0.9748	1.0000	0.9986	1.0000	1.0000	1.0000
10	25	0.0654	0.1204	0.0666	0.4734	0.1064	0.5102	0.1816	0.7976	0.2712	0.9138	0.3360	0.9552
	50	0.0810	0.2856	0.1202	0.8658	0.2522	0.9526	0.6022	0.9920	0.8646	1.0000	0.9522	1.0000
	75	0.1034	0.4154	0.2278	0.9836	0.4946	0.9972	0.9282	1.0000	0.9960	1.0000	1.0000	1.0000
	100	0.1230	0.5450	0.3660	1.0000	0.7360	1.0000	0.9954	1.0000	0.9980	1.0000	1.0000	1.0000
	125	0.1396	0.7264	0.5070	1.0000	0.8902	1.0000	0.9994	1.0000	0.9996	1.0000	1.0000	1.0000
15	25	0.0750	0.1762	0.0992	0.4898	0.1200	0.6716	0.2304	0.9192	0.2310	0.9828	0.4714	0.9938
	50	0.0956	0.3716	0.1860	0.9340	0.3276	0.9994	0.7592	1.0000	0.7594	1.0000	0.9920	1.0000
	75	0.1208	0.5854	0.3330	0.9972	0.6418	1.0000	0.9850	1.0000	0.9850	1.0000	1.0000	1.0000
	100	0.1508	0.7538	0.5168	1.0000	0.8828	1.0000	0.9996	1.0000	0.9998	1.0000	1.0000	1.0000
	125	0.1782	0.9086	0.6838	1.0000	0.9912	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	25	-	-	-	-	-	-	-	-	-	-	-	-
	50	0.1004	0.4450	0.2204	0.9890	0.4038	0.9986	0.8496	1.0000	0.8498	1.0000	0.9992	1.0000
	75	0.1450	0.7216	0.4224	1.0000	0.7728	1.0000	0.9804	1.0000	0.9974	1.0000	1.0000	1.0000
	100	0.1674	0.9080	0.5950	1.0000	0.9414	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	125	0.2202	0.9582	0.8168	1.0000	0.9970	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 2c: Size-Adjusted Power for the ADF-SUR and the DF-GLS-SUR Tests, $z_t = (1,t)$**,** no lag selection, size 5%

N	T	r = 0.99		0.97		0.95		0.9		0.85		0.8	
		ADF-SUR	DF-GLS-SUR	ADF-SUR	DF-GLS-SUR	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	0.2290	0.3490	0.3996	0.5748
	50	0.0510	0.0544	0.0616	0.0690	0.083	0.1012	0.2850	0.2906	0.4598	0.6318	0.8452	0.9064
	75	0.0514	0.0576	0.0830	0.0984	0.1412	0.1880	0.4960	0.6540	0.8896	0.9676	0.9952	0.9998
	100	0.0594	0.0638	0.0958	0.1418	0.2098	0.3330	0.7544	0.9144	0.9926	0.9986	1.0000	1.0000
	125	0.0596	0.0640	0.1400	0.1860	0.3412	0.4740	0.9486	0.9866	1.0000	1.0000	1.0000	1.0000
10	35	0.0542	0.0522	0.0550	0.0576	0.0762	0.0894	0.1518	0.2502	0.3516	0.5524	0.6372	0.8398
	50	0.0550	0.0524	0.0774	0.0874	0.1080	0.1452	0.3400	0.5238	0.7486	0.9178	0.9630	0.9964
	75	0.0556	0.0598	0.0964	0.1176	0.2050	0.2884	0.7604	0.9142	0.9952	0.9990	1.0000	1.0000
	100	0.0598	0.0644	0.1456	0.2002	0.3670	0.5412	0.9724	0.9966	1.0000	1.0000	1.0000	1.0000
	125	0.0668	0.0724	0.2058	0.3076	0.5768	0.7840	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15	35	0.0540	0.0548	0.0636	0.0732	0.0808	0.1150	0.1762	0.3654	0.4250	0.7356	0.7506	0.9606
	50	0.0538	0.0568	0.0748	0.0896	0.1204	0.1796	0.4466	0.6810	0.8728	0.9794	0.9416	1.0000
	75	0.0556	0.0588	0.1096	0.1530	0.2542	0.4002	0.8896	0.9838	1.0000	1.0000	1.0000	1.0000
	100	0.0588	0.0678	0.1656	0.2744	0.4608	0.7232	0.9966	1.0000	1.0000	1.0000	1.0000	1.0000
	125	0.0704	0.0844	0.2582	0.4304	0.7178	0.9100	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	35	-	-	-	-	-	-	-	-	-	-	-	-
	50	0.0508	0.0550	0.0688	0.0970	0.0084	0.2078	0.4596	0.7866	0.9104	0.9956	0.1000	1.0000
	75	0.0538	0.0586	0.1110	0.1778	0.2736	0.4850	0.9396	0.9960	1.0000	1.0000	1.0000	1.0000
	100	0.0648	0.0740	0.2008	0.3454	0.5688	0.8350	0.9990	1.0000	1.0000	1.0000	1.0000	1.0000
	125	0.0750	0.0920	0.3216	0.5276	0.8248	0.9738	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 2d: Size-Adjusted Power for the ADF-SUR and the DF-GLS-SUR Tests, $z_t = (1,t)$, with lag selection, size 5%

N	T	r = 0.99		0.97		0.95		0.9		0.85		0.8	
		ADF-SUR	DF-GLS-SUR	ADF-SUR	DF-GLS-SUR	ADF-SUR	DF-GLS-SUR	ADF-SUR	DF-GLS-SUR	ADF-SUR	DF-GLS-SUR	ADF-SUR	DF-GLS-SUR
5	35	0.0488	0.0498	0.0524	0.0576	0.0540	0.0790	0.0630	0.1596	0.0692	0.2938	0.0820	0.4546
	50	0.0492	0.0530	0.0556	0.0714	0.0676	0.1134	0.1214	0.3060	0.2074	0.5918	0.3204	0.7756
	75	0.0512	0.0552	0.0724	0.0980	0.1082	0.1932	0.2768	0.6234	0.5146	0.8838	0.7344	0.9658
	100	0.0578	0.0612	0.0938	0.1564	0.2854	0.3564	0.4902	0.8770	0.7802	0.9854	0.9310	0.9962
	125	0.0532	0.0662	0.1148	0.2164	0.2388	0.5332	0.6996	0.9694	0.9396	0.9970	0.9860	0.9994
10	35	0.0492	0.0514	0.0540	0.0620	0.0560	0.0828	0.0656	0.2226	0.0866	0.4398	0.1116	0.6506
	50	0.0522	0.0528	0.0678	0.0806	0.0888	0.1426	0.1776	0.4616	0.3254	0.7856	0.5200	0.9262
	75	0.0552	0.0570	0.0840	0.1274	0.1500	0.3124	0.4738	0.8600	0.7908	0.9882	0.9408	0.9992
	100	0.0588	0.0678	0.1160	0.2228	0.2590	0.5636	0.6460	0.9856	0.9718	0.9998	0.9976	1.0000
	125	0.0632	0.0766	0.1810	0.3864	0.4314	0.8360	0.9442	1.0000	0.9996	1.0000	0.9998	1.0000
15	35	0.0502	0.0516	0.051	0.0656	0.057	0.0944	0.0724	0.2610	0.0978	0.5284	0.1304	0.7502
	50	0.0500	0.0522	0.0706	0.0908	0.1028	0.1736	0.2426	0.5902	0.4510	0.7856	0.6872	0.9818
	75	0.0570	0.0622	0.0954	0.1668	0.1902	0.4128	0.5906	0.9508	0.9042	0.9884	0.9850	1.0000
	100	0.0592	0.0656	0.1362	0.3090	0.3264	0.7452	0.8952	0.9984	0.9962	1.0000	1.0000	1.0000
	125	0.0648	0.0858	0.2208	0.4956	0.5534	0.9364	0.9904	1.0000	1.0000	1.0000	1.0000	1.0000
20	35	-	-	-	-	-	-	-	-	-	-	-	-
	50	0.0516	0.0560	0.0704	0.0998	0.0978	0.1972	0.2606	0.6760	0.8550	0.9428	0.7438	0.9952
	75	0.0574	0.0656	0.1098	0.1806	0.2198	0.4744	0.6836	0.9816	0.9560	0.9996	0.9968	1.0000
	100	0.0648	0.0730	0.1712	0.3558	0.4192	0.8272	0.9612	1.0000	1.0000	1.0000	1.0000	1.0000
	125	0.0732	0.0960	0.2536	0.5914	0.6512	0.9748	0.9988	1.0000	1.0000	1.0000	1.0000	1.0000

Table 3: Univariate Unit Root tests

	ADF			DF-GLS		
	a	t_a	k^{GS}	a	t_a	k^{MAIC}
Australia	-0.3014	-0.3014	0	-0.0055	-0.2447	1
Austria	-0.0638	-2.0407	0	-0.0343	-1.3519	6
Belgium	-0.0440	-1.5798	0	-0.0538	-1.9745**	1
Canada	-0.0083	-0.6009	4	0.0037	0.3204	4
Denmark	-0.0421	-1.4943	0	-0.0418	-1.7438*	1
Finland	-0.0541	-1.7942	0	-0.0341	-1.4104	0
France	-0.0549	-1.7576	0	-0.0720	-2.3718**	1
Germany	-0.0549	-1.7720	0	-0.0699	-2.0650**	6
Greece	-0.0628	-1.9407	0	-0.0649	-2.1180**	5
Ireland	-0.0946	-2.4182	0	-0.0658	-1.9506**	0
Italy	-0.0647	-1.9340	0	-0.0646	-1.9398*	0
Japan	-0.0442	-1.8689	5	-0.0138	-0.9673	1
Netherlands	-0.0585	-1.8615	0	-0.0603	-2.0685**	1
Norway	-0.0419	-1.3657	0	-0.0442	-1.7194	1
New Zealand	-0.0271	-0.9547	0	-0.0491	-1.7777	1
Portugal	-0.0494	-1.7244	0	-0.0366	-1.4140	0
Spain	-0.0508	-1.8404	0	-0.0360	-1.7018	2
Sweden	-0.0066	-0.2838	0	-0.0168	-0.7311	1
Switzerland	-0.0657	-2.1218	0	-0.0273	-1.4288	1
U.K	-0.0889	-2.2337	5	-0.0471	-1.6549	0

Critical Values are from Mac Kinnon (1991), adjusted to 73 observations, for the ADF test, and from Hamilton for DF-GLS. *, **, *** represent reject at 10%, 5% and 1%.

The ADF regression is $\Delta q_t = m + a q_{t-1} + \sum_{i=1}^{k^{GS}} y_i \Delta q_{t-i} + u_t$

The DF-GLS regression is $\Delta q_t^m = a q_{t-1}^m + \sum_{i=1}^{k^{MAIC}} y_i \Delta q_{t-i}^m + u_t$

Figure1. Robustness Analysis

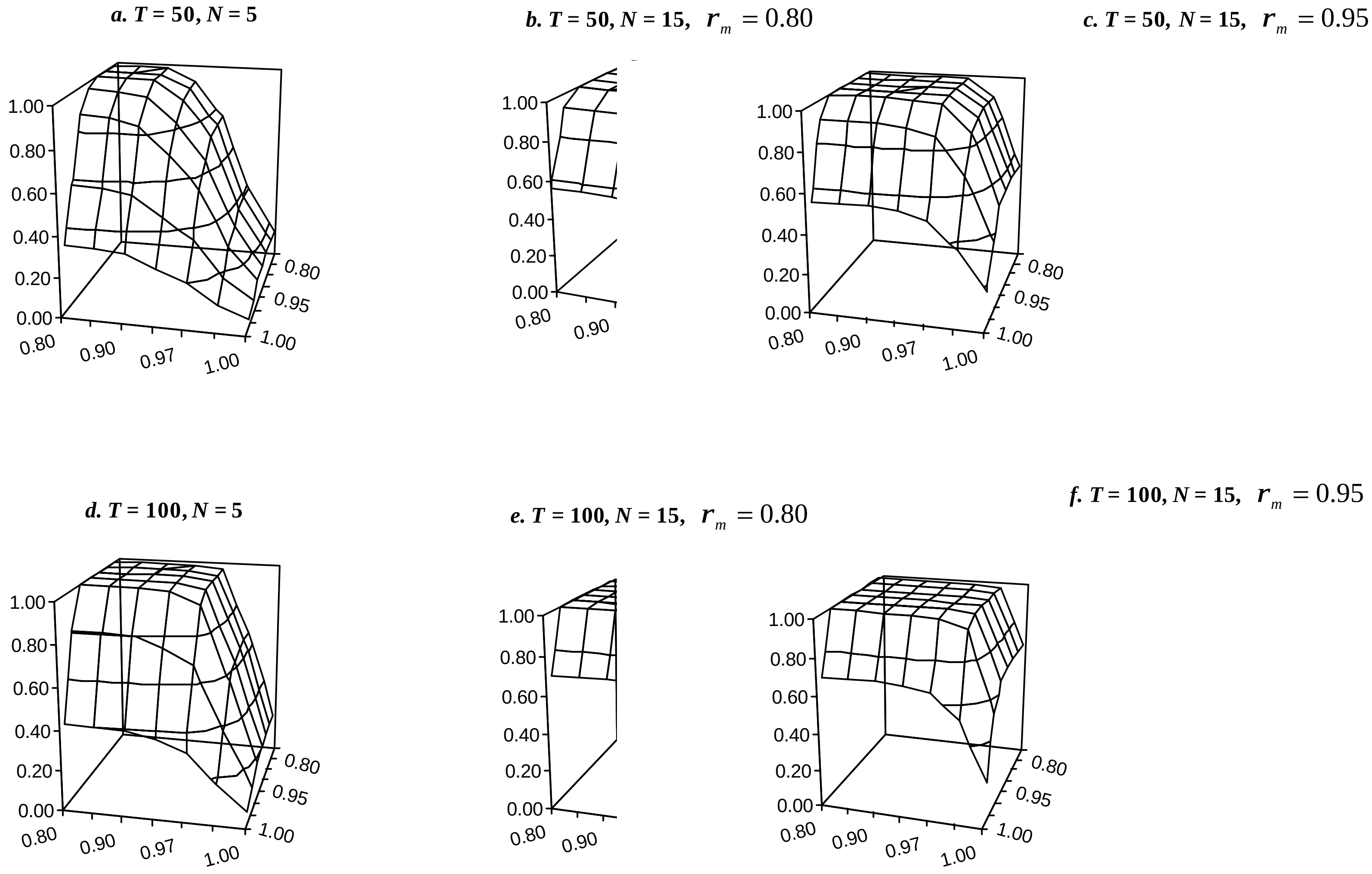

Figure2. The IPS Test v.s. the DF-GLS-SUR Test

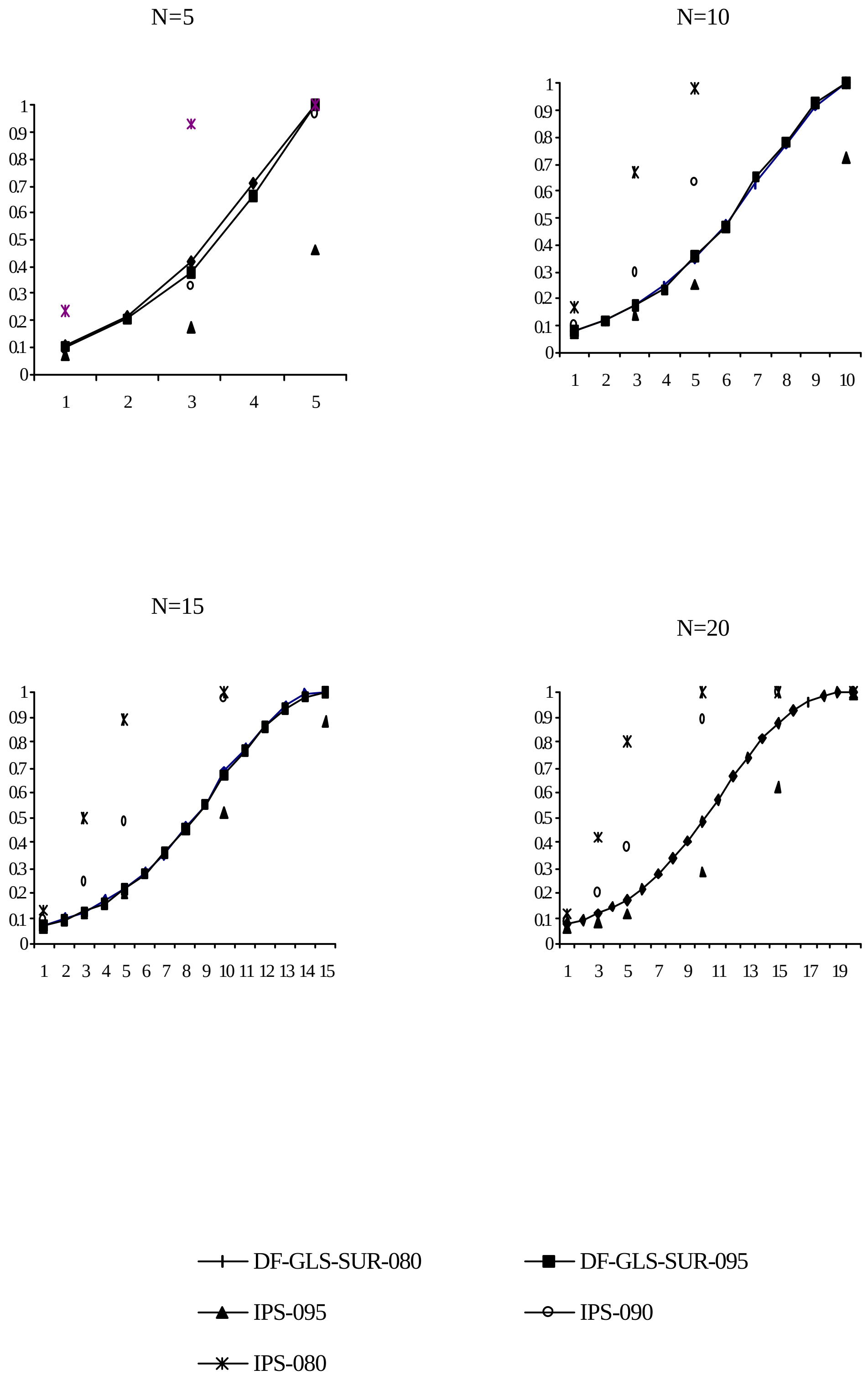

Figure 3: Multivariate Unit Root Tests, P-values