The Quest for Purchasing Power Parity with a Series-specific Unit Root Test using Panel Data

Janice Boucher Breuer
Department of Economics
The Darla Moore School of Business
University of South Carolina
Columbia, SC 29208
email:boucher@darla.badm.sc.edu

Robert McNown
Department of Economics
Campus Box 256
University of Colorado
Boulder, CO 80309
email:robert.mcnown@colorado.edu

Myles Wallace
Department of Economics
Sirrine Hall
Clemson University
Clemson, SC 29634-1309
email:myles@clemson.edu

December 2000

*We wish to thank Donggyu Sul for comments on an earlier version of this paper.

Clemson University Center for Internation Trade
CUCIT Working Paper

Abstract

A unit root testing procedure is presented that exploits the well-established power advantages of panel estimation while rectifying a deficiency in other panel unit root tests. This procedure, which takes into account contemporaneous cross-correlation and heterogeneous serial correlation of the regression residuals, allows determination of which members of the panel reject the null hypothesis of a unit root and which do not. Applying the procedure to real exchange rates yields results that are in broad agreement with those obtained from single-equation unit root tests. There is little evidence that a unit root can be rejected in dollar-based real exchange rates for the floating rate period.

I. Introduction

Prior to 1996, evidence in support of purchasing power parity for the floating exchange rate period for dollar-based currencies was rare. Single-equation augmented Dickey-Fuller tests for a unit root in real exchange rates, which became the tool for such investigations, failed to produce much support for purchasing power parity. (For a survey, see Breuer (1994)). The common finding across such studies was that "real exchange rates contain a unit root." Occasionally, a study would find evidence rejecting the null hypothesis of a unit root for a particular currency, but intermittent findings like these were, perhaps justifiably, regarded as anomalies in the vast pre-1996 empirical literature that had systematically found little support for purchasing power parity over the floating rate era.

Post-1996 evidence offers a different conclusion. A revival of empirical support for purchasing power parity has occurred with the application of panel unit root estimation procedures to testing for a unit root in real exchange rates. Studies by Sarno and Taylor (1998), Taylor and Sarno (1998), Wu and Wu (1998), Papell (1997), Oh (1996), Wu (1996), and others, using a cross-section of real exchange rates, present evidence for the floating rate period that favors stationarity. Real exchange rates are found to be mean reverting in these studies, and so purchasing power parity is corroborated. O'Connell (1998), Papell[1] (1997) and Papell and Theodoridis (1998) provide dissenting evidence in their panel unit root tests.

The evidence of mean reversion is often bittersweet for proponents of purchasing power parity; the half-life estimates for shocks to the real exchange rate range between two to four years, suggesting considerable persistence in the departures from purchasing power parity. The conclusions from panel studies should also be regarded as only weakly supportive of purchasing power parity since the extant panel studies (including tests presented by Im, Pesaran, and Shin (1997), Maddala and Wu (1997), and Taylor and Sarno's (1998) MADF test) are tests of a joint hypothesis that all real exchange rates in the panel contain a unit root. When rejection of the joint null hypothesis materializes, it is possible that only one member of the panel contributes to the finding thus making any proclamations favorable to purchasing power parity less strong.

This paper presents a panel estimation procedure that can be used to help identify the time series properties of *each* individual member in a panel. Thus, the procedure allows a researcher to avoid the joint null hypothesis problem endemic to extant panel unit root tests while still taking advantage of contemporaneous cross-correlations of the regression residuals and increased sample size afforded through panel estimation. The procedure also handles heterogeneous serial correlation across panel members. Using this estimation strategy, we find very little evidence supportive of purchasing power parity; our results echo findings from single-equation unit root tests and present a stark contrast to the results from many recent panel unit root studies.

To set the stage for the procedure we use, the next section contains a brief review of some current panel tests.[2]

II. A Brief Review of Some Panel Unit Root Tests

Panel unit root studies of real exchange rates come on the heels of the development of statistical approaches to conducting unit root tests with cross-sectional and time series data (panel data) formally introduced by Quah (1990), Breuting and Meyer (1991), and Levin and Lin (1992, 1993). However, the proper genesis of panel unit root tests predates the works of Levin and Lin and Quah, and actually begins with Abuaf and Jorion (1990) who introduce a panel unit root estimation procedure to test purchasing power parity.[3] Their motivation, as with subsequent panel unit root studies, is a desire to increase the power properties of single-equation unit root tests that are based on a limited time series dimension.

The general model for N currencies and T time periods that encompasses all panel unit root tests is:

k1

$$\begin{array}{l} \Delta q_{1,t} = \alpha_1 + \beta_1 \cdot q_{1,t-1} + \sum_{j=1} \delta_{1,j} \cdot \Delta q_{1,t-j} + u_{1,t} \quad t = 1,...,T; \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ \Delta q_{N,t} = \alpha_N + \beta_N \cdot q_{N,t-1} + \sum_{j=1}^{kN} \delta_{N,j} \cdot \Delta q_{N,t-j} + u_{N,t} \quad t = 1,...,T; \end{array} \tag{1}$$

where $q_{i,t}$ is the real exchange rate for currency i at time t. In their early application of this model to test PPP, Abuaf and Jorion restricted all β_i's to be identical, and omitted the lagged differences from each equation. The unit root null hypothesis ($\beta_i = \beta = 0$, i=1,2,...,N) and the stationary alternative apply to all N currencies jointly. In their application to the G-10 counties for the 1973-1987 period, Abuaf and Jorion reject the unit root hypothesis and conclude that all currencies in the panel are stationary.

Levin and Lin (1992, 1993) provide statistical foundations for the panel unit root tests, and generated tables of critical values for a range of values of N and T in models with either fixed effects or a single intercept restricted to be identical across the panel. While lagged augmentation terms could be added to their specification to correct for serial correlation of the error terms, no allowance is made for contemporaneous cross-correlation of the errors. Levin and Lin also retain the restrictive joint null and alternative hypotheses of the Abuaf and Jorion approach, since the β_i are restricted to be identical across the panel under both hypotheses.

Studies by Wu (1996) and Oh (1996) use variants of the Levin and Lin procedure to test for a unit root in real exchange rates. While both studies find evidence of PPP for the recent float, these studies impose a homogeneous lag structure across panel members and they do not account for contemporaneous cross correlation of the error terms. If the error terms are contemporaneously cross-correlated, the Levin and Lin critical values used in these studies will not be appropriate, and the conclusions regarding purchasing power parity could be erroneous.

Papell (1997) adapts the Levin and Lin specification to allow for heterogenous serial correlation of the errors in testing for purchasing power parity.[4] The lag structure is permitted to differ across the panel members, but possible contemporaneous error correlations are not considered. Papell finds evidence favorable to purchasing power parity during the floating rate era for a panel of twenty countries with monthly (but not quarterly) data regardless of the choice of the price index used to construct real exchange rates.

O'Connell (1998) adapts the LL specification to test for a unit root in real exchange rates by taking into account contemporaneous cross-sectional dependence in the data. O'Connell also incorporates lags of $\Delta q_{i,t}$ to account for potential serial correlation but forces the lag structure across panel members to be homogeneous so that the δs (and k's) are identical. He also simulates critical values that are based on cross-sectional dependence. O'Connell finds that once cross-sectional dependence is accounted for, a unit root in real exchange rates cannot be rejected for panels of up to 64 countries.

A serious problem with all adaptations of the Levin and Lin test discussed above is that all panel members are forced to have identical orders of integration. The null hypothesis is that all series contain a unit root, while the alternative hypothesis is that no series contains a unit root, i.e. all are stationary. Breuer, McNown, and Wallace (1998) demonstrate with Monte Carlo simulations how the Levin and Lin procedure performs in panels with mixed orders of integration. With as few as one stationary member of the panel, the rejection rate rises above the nominal size of the test, and continues to increase with the number of I(0) series in the panel. Although the null hypothesis that all series have a unit root is correctly rejected, the alternative of "all stationary" is also false in these mixed panels. Consequently, rejection of the null hypothesis with these LL-type tests may have led to overinflated support of purchasing power parity.

Recognizing this problem, recent papers by Im, Pesaran and Shin (1997), Maddala and Wu (1997), Sarno and Taylor (1998), Taylor and Sarno (1998), and Wu and Wu (1998) present second generation panel unit root tests that allow the autoregressive coefficient to differ across panel members under the alternative hypothesis. In contrast to the LL-type tests, rejection of the null hypothesis with these tests means that "not all members of the panel contain a unit root." Thus, these tests all admit that there may be a mixture of stationary and nonstationary processes in the panel under the alternative hypothesis.

Im, Pesaran, and Shin's (IPS) test is constructed as a simple average of the t-statistics on the autoregressive coefficient generated from N single-equation augmented Dickey-Fuller tests. Because the IPS statistic is based on N independent augmented Dickey-Fuller equations, it permits different autoregressive coefficients as well as heterogeneity of lag structures in the N individual series, but does not account for contemporaneous cross correlation.[5]

Maddala and Wu (1997) use single-equation OLS estimation similar to Im, Pesaran, and Shin except that the p-values corresponding to the individual t-statistics are used to construct the (Fisher) test statistic. Wu and Wu (1998) modify both the IPS and Fisher tests to take into account the contemporaneous correlation of the residuals by using SUR estimation. They report evidence that is generally favorable to PPP during the recent float.

Sarno and Taylor (1998) and Taylor and Sarno (1998) propose two tests that also allow the autoregressive coefficient across panel members to be different. The first test Sarno and Taylor propose is the multivariate augmented Dickey-Fuller (MADF). This test is based on seemingly unrelated regressions (SUR) estimation of the unrestricted version of model (1), allowing both heterogeneous lags structures and contemporaneous error correlations across the panel. Under the alternative hypothesis, the β coefficients are permitted to differ across the panel members. The MADF test is a joint test of the null hypothesis of $\beta_1 = \beta_2 = ... = \beta_N = 0$. Taylor and Sarno demonstrate that this test has power properties that are significantly better than the single equation augmented Dickey-Fuller test. Additionally, the power increases as the number of panel members that are defined as stationary in the simulation increases.

A problem with the tests proposed by Im, Pesaran, and Shin (1997), Maddala and Wu (1997), Sarno and Taylor (1998) and Taylor and Sarno (1998) stems from the joint hypothesis that is tested. Under the joint hypothesis, rejection does not provide information about how many panel members reject the null hypothesis and how many do not. Furthermore, if there is a mixture of unit root and stationary processes, none of these tests indicate which panel members are stationary and which are not.

Acknowledging this problem, Sarno and Taylor (1998) and Taylor and Sarno (1998) present a second test, the JLR test, which is complementary to the MADF test. The JLR test is an application of the Johansen-Juselius (1990) test for cointegration in a system of N variables and is based on the rank of the long run impact matrix Γ_k (the coefficient matrix on the lagged levels) in a vector error correction representation of this system. The null hypothesis is that Γ_k is less than full rank, implying at least one unit root, versus the alternative hypothesis that Γ_k is of full rank. Rejection of the null implies that all series are stationary.

When used together the MADF and JLR tests provide some limited information on the mix of stationary and nonstationary series in a panel. For example, if the MADF rejects that all series have a unit root, but the JLR test does not reject that the rank of Γ_k is less than full, then a reasonable inference is that at least some of the series have a unit root. Unfortunately, in this case the proportions of I(0) and I(1) series remain unknown. In addition, because the JLR is derived from an error correction framework degrees of freedom are rapidly used up as the panel expands in size and the test cannot be used in large panels.

III. Individual Tests of a Unit Root in Real Exchange Rates in a Panel Setting

This section outlines a panel unit root estimation procedure and test statistics that can be used to identify which real exchange rates in a panel contain a unit root and which ones do not. This test incorporates advances in previous panel tests by allowing heterogeneous serial correlation across the panel, contemporaneous correlation among the errors, and different autoregressive parameters for each panel member under the alternative hypothesis. The innovation here is that separate null and alternative hypotheses are tested for *each* panel member within a SUR framework. Although this structure of hypotheses is the same as for single-equation unit root testing, the SUR model exploits the information in the error covariances to produce efficient estimators and potentially more powerful test statistics.

As in other unit root tests, the test statistics from the SUR model have nonstandard distributions with critical values that must be derived by simulation. The simulations must produce critical values appropriate for testing the null hypothesis that $\beta_i = 0$ for each individual member of the panel. The computed critical values will be specific to the estimated covariance matrix for the system of real exchange rates considered and the sample size and number of panel members. Consequently, the critical values reported here are specific to this study and cannot be applied to other data sets.

The test presented here has features of both single-equation and panel estimation in tests for a unit root. There are several advantages to this procedure. First, like the predecessor work, this test exploits the improved power properties of moving from single-equation to panel unit root tests. Second, the test statistics are produced by estimators that incorporate cross equation residual covariances and allow heterogeneity in the form of serial correlation across the panel members. Third, the procedure allows identification of how many and which members of the panel contain a unit root and which do not. The last advantage arises because the framework tests individual hypotheses, rather than joint hypotheses as in previous versions of the panel unit root tests.

Some readers would argue that it is inappropriate to examine the possibility that purchasing power parity holds for some exchange rates and not others. However, this all-or-none view of purchasing power parity is inconsistent with both theoretical considerations and empirical evidence from a range of countries and time periods. If monetary shocks dominate, the results may be more favorable to purchasing power parity than when real shocks to consumption or production dominate. *A priori*, there is no reason to believe that the nature of shocks is the same for countries. For example, evidence from high inflation countries is generally more favorable to purchasing power parity than is that from industrialized countries during the recent float (McNown and Wallace, 1989).

III.A Data

To compare test results with those reported previously, this study examines three panels of real exchange rates that have been investigated by others. The three panels consist of: (1) the United Kingdom, France, Germany, and Japan, as examined by Taylor and Sarno (1998); (2) the EMS -7 (Belgium, Denmark, France, Germany, Ireland, Italy, and the Netherlands) tested by Papell (1998) and Wu and Wu (1998); and (3) the industrial 20 (EMS-7 plus Canada, Australia, Japan, United Kingdom, Switzerland, Austria, Finland, Greece, Norway, Portugal, Spain, Sweden, and New Zealand) also studied by Papell (1997) and Wu and Wu (1998).

Quarterly data on nominal exchange rates and consumer price indices of countries are gathered from the IFS CD-ROM version 12/96 and the IFS 1998 yearbook so that the sample spans 1973:2 - 1998:1. Real exchange rates equal the logarithm of the end-of-period nominal exchange rates (line rf) in foreign currency per U.S. dollar times the U.S. consumer price index divided by the foreign consumer price index (base year=1990, line 64).

III.B Single-equation Tests for a Unit Root in Real Exchange Rates and Residual Correlations

Table 1 presents the results for the individual Augmented Dickey-Fuller tests (ADF). After adjusting for differencing and lags ninety-one observations are available for the dependent variable spanning the period 1975:3 - 1998:1. Starting with a maximum of eight lags and testing down, lags were eliminated when insignificant using a standard t-test and a 0.10 level of significance. Based on Q-statistics lags are included, even when insignificant, if their deletion resulted in serial correlation. Table 1 reports very limited support for PPP. Only at the 0.10 significance level can a unit root be rejected, and then for only seven out of twenty countries.

The single-equation ADF tests do not take advantage of the possibility that the errors across the various exchange rates might be correlated. If the residuals are correlated, the efficiency of SUR over independent OLS increases as the average residual correlation rises. If the gain in efficiency translates into increased power, testing within the SUR framework may produce evidence more favorable to mean reversion among real exchange rates. As reported in Table 3, the average of the absolute value of the correlations of each series' residuals with the residuals from all others series in the panel are substantial. Correlations of the ADF residuals for the panels of four and seven countries are in excess of 0.60 and are higher on average for the panel of seven EMS countries than for the four considered by Taylor and Sarno. For the twenty industrial countries, the correlations are all in excess of 0.25 except for Canada which was an outlier with an average correlation of 0.06. In fact, out of 31 possible correlations 28 exceed 0.5. These high residual correlations are indicative of the potential gain in information when using panel estimation and unit root tests compared to single-equation tests.

III.C. Panel Tests for a Unit Root in Real Exchange Rates

To take into account the cross equation correlations between the residuals, all equations were re-estimated using SUR, while allowing all parameters to differ across the panel. Three groupings were estimated as mentioned above: the Sarno and Taylor (1998) four; the EMS-7; and finally, a more inclusive group of 20 industrial countries. The numbers of lagged augmentation terms for each series were set equal to those found for the individual ADF tests. The individual country t-statistics on the lagged level of the real exchange rate from the SUR estimation for the various groupings are found in Table 2 under the column SURADF.

The Multivariate Augmented Dickey-Fuller test (MADF) first implemented by Sarno and Taylor

(1998) is also reported in Table 2. For each panel of countries, the MADF statistic tests the hypothesis that the coefficients on the lagged level of each real exchange rate are *jointly* equal to zero. Nonrejection of the null implies that all real exchange rates have a unit root. Rejection of the unit root null in panel tests such as the MADF implies that at least one real exchange rate is stationary. However, the test provides no information about the number of stationary members in the panel, or which ones are stationary.

Note that the ADF statistics from the SUR regressions reported in Table 2 (subsequently called SURADF statistics) are generally substantially larger in absolute value than those found in the single equation tests presented in Table 1. Use of the standard Dickey-Fuller critical values would result in a rejection of a unit root for most of the countries. However, reliance on these critical values would now be inappropriate since a panel environment is used to estimate and test $\beta_i = 0$. Therefore, to confront the null hypothesis of a unit root in environments where panel estimation has been used with nonstationary data, critical values specific to the information contained in the estimation must be generated.

The simulation of critical values was based on environments that matched the parameter estimates from the SUR estimation for each group of countries. In particular, error series were generated to be normally distributed with variance covariance matrix given by the SUR estimates for each panel. Then each simulated real exchange rate was generated from the error series, using the SUR estimated coefficients on the lagged differences. Since the null hypothesis is a unit root without drift, both the lagged level and the intercept were set equal to zero.[6]

The number of observations for each series in the simulation was set to 141. To minimize the sensitivity of the results to initial conditions (the initial values for each simulated real exchange rate were zero) the first 50 observations on the dependent variable were omitted. This left 91 usable observations for each series, corresponding exactly to the number of observations on the dependent variable in the real data. The number of replications was set at 10,000. Results for the 0.01, 0.05, and 0.10 critical values applicable to each country based on each panel, and for the MADF test for each panel, are displayed in

Table 2.[7]

The SURADF critical values for each panel of countries reported in Table 2 are substantially higher than those applicable to the single equation ADF test. Although most computed SURADF test statistics are larger in absolute value than the ADF statistics, they rarely exceed the larger critical values now appropriate for the SUR environment. Based on the SURADF critical values, evidence favorable to purchasing power parity at the 0.10 significance level or better is found for three of the four countries in the four-country panel of Taylor and Sarno (1998). This conclusion is slightly different from Taylor and Sarno's conclusion that all four exchange rates in this panel are stationary.

From the other two panels there is virtually no support for purchasing power parity. No rejections are reported for any country in the EMS-7 panel and in the industrial twenty panel, the only significant SURADF test statistics are for Finland (at .05 level) and New Zealand (.10 level). These results are consistent with those generally obtained from the pre-panel, single-equation studies, the single equation ADF tests we report in Table 1, and the more recent panel findings reported by O'Connell (1998) and Papell (1998) for quarterly data.[8, 9]

The MADF statistic tells a similar story. The null hypothesis that all series are unit root processes is rejected at 0.10 significance level only for the Sarno and Taylor group of countries. However, it would be incorrect to infer from the MADF that all real exchanges rates in the Sarno and Taylor grouping are stationary, since evidence from the SURADF test indicates that the UK real exchange rate is a unit root process.

Of course, if the SURADF test has little power to reject a false null of a unit root for individual series that are actually stationary, the evidence from the larger seven and twenty country panels in Table 2 that is unfavorable to PPP may be misleading. However, evidence will be presented in section IV that the SURADF test generally has substantially higher power than the single equation ADF test for a sample size of 100 or greater. This supports the conclusion that the evidence in Table 2 provides little support for purchasing power parity.

III.D. Discussion of Critical Values

The efficiency of SUR estimation relative to OLS increases with the average absolute size of the error correlations and decreases with the average absolute magnitudes of the correlations among the regressors across equations (Greene, 1997). When each equation in the SUR system is an autoregression as in the case of panel unit root testing, then the average correlation among the regressors is positively related to the corresponding average error correlation. In the simple case of an autoregressive equation of order one with each equation having the same AR coefficient, the correlations among the errors are identical to those between the regressors.[10] In the panels estimated in this study the lag specifications differ across the equations, so that the relation between error correlations and correlations among the regressors, and the implications for SUR efficiency gains, is complex.

It is apparent from Table 2 that the critical values increase in absolute value with the number of members (N) in the panel. Consider the results for the 0.05 critical values. Disregarding the minus sign, for the four country panel all critical values are less than four while the critical values for the seven country EMS panel are all greater than four, but none are as large as five. However, in the panel of twenty, twelve critical values exceed five.[11] Also, as seen in Table 3 as the average correlation between residuals from each equation and those from the other panel members increases, so does the absolute critical value for that specific equation.

These two observations are investigated more formally by regressing the 0.05 critical values for each series found in Table 3 (31 observations) on the number of countries in each panel and the average correlation of the own series residuals with the residuals from all other series in the panel.[12] Standard errors of the coefficient estimates are reported in parentheses.

$$CV05 = -0.12 - 4.24 \cdot AVECORR - 0.13 \cdot NUMCOUNTRY$$
$$\quad\quad (0.24) \quad (0.50) \quad\quad (0.01)$$
$$\text{adjusted } R^2 = 0.79$$

The regression explains nearly 80% of the variation in the critical values, with the critical values "increasing" (i.e. becoming more negative) with the size of the panel and the average correlation of the residuals.

Additional critical values for various time series dimensions, T, are found in Table 4. The procedure was identical to the one used to generate the appropriate critical values for a sample size of 91. In this case, as expected, the absolute critical values decrease as the sample size T increases with most of the decrease occurring between T = 25 and T = 100.[13] Of course, these critical values are only applicable for the lag structure and covariance matrix used for these simulated panels. Copies of the program used to generate research-specific critical values is written in RATS and is freely available on request.

IV. Power Analysis of Individual Unit Root Tests in Panel Settings

Since the evidence from the EMS and twenty-country panels was generally unfavorable to purchasing power parity, it is important to determine if this outcome reflects a low power of the SURADF test. To relate the power properties of various unit root test statistics to the empirical environment, the lag structure of each of the series and the error covariance matrix are set equal to the corresponding SUR estimates. The power of all three tests (ADF, SURADF, and MADF) is investigated in experiments involving 5,000 replications in two alternative sets of environments. In all cases the size of the test is equal to 0.05, testing against critical values reported in Table 4.

In the first set of environments, every series in the panel is simulated to be a stationary, I(0) series with common autoregressive coefficients of 0.99, 0.95, or 0.90. The panels consist of four, seven, or twenty countries, and samples size of T = 25, 100, 250, and 500.[14] The second set of environments involves a twenty country panel that mixes stationary and nonstationary processes, estimated with a sample size of 100. The stationary series each have an autoregressive coefficient of 0.99, 0.95, or 0.90. Series that are stationary are generated with non-zero intercepts equal to those obtained from the SUR regressions.[15, 16]

IV.A. Power Analysis under the Alternative Hypothesis that all Series are I(0)

Tables 5-7 report the rejection rates for the ADF, SURADF, and MADF tests in the first set of environments in which all members in the panels are generated as stationary series. Table 5 reports the results for a panel of size four, Table 6 for a panel size of seven, and Table 7 for a panel size of twenty. The 0.05 critical values from Table 4 for the ADF and SURADF tests (0.95 for MADF), corresponding to the panel size (N), sample size of T=100, and based on the null that all series in the panel are I(1), are used in the power analysis. Since the null hypotheses of the ADF and SURADF tests are of a unit root in each panel member while the null hypothesis of the MADF test is that all panel members contain a unit root, a comparison of power between the MADF and the other two tests is inappropriate. The analysis reported here focuses on the relative power of the ADF and SURADF, which have the same null and alternative hypotheses. However, we present independent information on the power for the MADF test, to contribute to the work of Sarno and Taylor (1998) who considered a smaller set of environments in their investigation of MADF power.

Table 5 shows that for the four-country panel of Sarno and Taylor at the smallest sample size and the largest value of the autoregressive parameter, both the SURADF and the ADF tests have limited power to reject a false null hypothesis. There is also little difference in power when the autoregressive coefficient is 0.95 or 0.90 and the sample size is 500, or when the autoregressive coefficient is 0.90 and the sample size is 250. In these cases the power approaches 1.0 for both tests. However, the SURADF test is often substantially more powerful than the ADF test for intermediate sample sizes and smaller values of the autoregressive parameter. Focusing on the case of T=100 and ρ=0.95 as representative of recent PPP tests with quarterly data, the gains in power across the four countries range from 37% to 100%. The SURADF test also has considerably more power than the ADF test when the autoregressive parameter is very close to one (0.99) with the largest sample size (T=500). The SURADF rejection rates range between 25% and 45% in this environment.

Table 6 reports the rejection rates for the panel consisting of the seven EMS countries. For brevity, Table 7 reports the lowest (L), median (M), and highest (H) rejection rates from the panel of twenty members. For example, in Table 7 when ρ = 0.95 and T = 100, among all twenty series in the panel the lowest rejection rate for an individual series using the SURADF is 15.2%, the highest is 48.1%, and the median value 24.8%.

Tables 6 and 7 display the same general pattern as in Table 5. At one extreme (an autoregressive root of 0.99 and a small number of time series observations) the power of the ADF and SURADF tests are low, and at the other extreme with a large sample size and smaller autoregressive parameter the power for both tests approaches 1.0. However, for all other cases the gain in power of the SURADF test over the ADF test is substantial, generally ranging in magnitude between a factor of two to almost four.

Overall, the results in Tables 5-7 indicate that the SURADF test has substantially more power than the single equation ADF test to reject a false null of a unit root when intermediate sample sizes are combined with moderate values of the autoregressive coefficient. In addition, there are substantial gains in power even when the autoregressive coefficient is 0.99 as long as the sample size is large (250 or 500). Since studies of purchasing power parity published in the 1990s typically work with sample sizes close to 100 observations (with quarterly data) and report autoregressive coefficients in the neighborhood of 0.95,[17] one can expect double to triple the power with the SURADF test as compared with the ADF test in this environment. Furthermore, since the SURADF test exhibits moderately high power in this case, the general lack of support for stationary real exchange rates shown in Table 2 cannot be criticized as due entirely to "poor power performance."

A comparison of SURADF power across the panels in Tables 5-7 shows that moving to larger panels does not substantially alter the power. In general, the tables suggest that for our panels, panel size has a relatively small effect on SURADF power.

Finally, the MADF test exhibits rejection frequencies that are often, but not always, higher than the mean rejection rates of the SURADF test. However, since the null and alternative hypotheses differ for these two tests, a direct comparison of these results is not meaningful. Furthermore, rejection of the null hypothesis with the MADF test can only indicate that at least one series in the panel is stationary, offering no information about the number of series that are stationary, or which ones. This issue is analyzed further in the next section.

IV.B. Power Analysis under the Alternative Hypothesis that not all series are I(0)

In most cases the researcher will not know the mix of stationary and non-stationary series in the panel. For this reason this section briefly explores the power of the SURADF test in mixed I(0) and I(1) samples.

As in previous simulations the autoregressive coefficient for all of the I(0) members is set to 0.99, 0.95, or 0.90. All series are generated with the lag structure and error covariances found in the SUR estimations. Intercepts are set equal to zero (no drift) for the I(1) series and are set equal to their estimated mean values for the I(0) series. The power analysis for this mixed environment is based on a sample size of 100 observations and 5,000 replications. The 0.05 critical values found in Table 4 for the twenty industrial countries and a sample size of T=100 under the null hypothesis that all series are I(1) are used in the power studies.

Using the large panel of twenty countries three specifications are investigated with results presented in Tables 8-10. The first sets the four Sarno and Taylor countries to be I(0) with the remaining sixteen constructed as I(1). The second sets the seven EMS countries to be I(0) with the remaining thirteen constructed as I(1) series. The third reverses the setup for the EMS countries. The seven EMS countries are now constructed to be I(1) with the remaining thirteen countries I(0). These environments differ from those in Tables 5-7 in two respects. Panel size (and therefore panel membership) is constant at twenty panel members and for the power analysis not all members are I(0).

The power results for the mixed panel containing four I(0) and sixteen I(1) series are presented in Table 8. Also reported are the rejection rates for the remaining sixteen unit root series -- these are the rejection rates of the null hypothesis when it is true and are referred to as the empirical size of the test.[18] For these sixteen I(1) series we report the lowest, median, and highest rejection rates.

Except for values of the autoregressive coefficient equal to 0.99, the SURADF test is (with one exception) a more powerful test than the ADF, sometimes substantially so. When the autoregressive coefficient is either 0.95 or 0.90 the SURADF test has about double the power of the ADF test for the UK, and more than triple the power for France and Germany.

The results from Tables 9 and 10, with seven and thirteen stationary series, respectively, are also favorable to the SURADF test. Based on median rejection rates the SURADF test has double (Table 10) to triple (Table 9) the power of the single equation ADF test to reject the null hypothesis when the autoregressive coefficient on each I(0) series is 0.90. The power gains are less dramatic for an autoregressive coefficient of 0.95 in Table 10, but are overwhelmingly favorable to the SURADF test for the different environments reported in Table 9, ranging from about double up to a factor of seven. When the autoregressive coefficient is 0.99, the ADF and SURADF tests both have comparable, but low power to reject a false null. Again, these results suggest that the failure to reject a unit root in real exchange rates with the SURADF test is not wholly due to a lack of power.

One purpose of this mixed panel power analysis is to reconcile the differences in empirical purchasing power parity results across the three panels reported in Table 2. The SURADF test presents evidence favorable to purchasing power parity for the Sarno and Taylor panel of four countries, but does not reject a unit root for these same countries when included in a larger panel of twenty industrial countries, or for Germany and France in the seven-country panel. The power investigations demonstrate the SURADF test has greater power than the ADF test, and shows little evidence of reduced power in mixed samples. Thus, the empirical results cannot be attributed solely to low power in the larger panel environments. With power less than 100%, type II errors can occur, and inferences may differ across panels for any given realization of the data. The MADF shows similar inconsistency, rejecting a unit root for all series in the smaller panel of four at 0.10, but not for the other two larger panels in Table 2. If we use 10% critical values instead, the results from the four country and twenty country panels do not differ as sharply as first appears. Table 2 displays SURADF statistics for Japan, Germany, and France that are close to their 10% critical values in the twenty-country panel, and the MADF statistic comes close to its 10% critical value in the largest panel as well.

To summarize, the simulations indicate that the SURADF test is in nearly all cases more powerful than the ADF test. In environments with moderate sample sizes and smaller values of the autoregressive parameter, these differences are substantial, with the SURADF exhibiting double to as high as seven times the power of the ADF. However, the SURADF test is affected by panel composition and size, so that its power will vary across different panels. The complete set of power experiments demonstrates that for a given panel size of N, the rejection frequency for a false unit root hypothesis depends upon (i) the number of time series observations, (ii) the value of the largest autoregressive coefficient in the process, (iii) the extent of cross-correlation between the errors; (iv) the extent of cross-correlation between the regressors, and (v) the mix of I(0) and I(1) series in the panel.

V. Conclusion

Purchasing power parity in real exchange rates is examined across three panels of real exchange rates for the floating rate period using the augmented Dickey-Fuller unit root test estimated in a SUR environment. The SURADF test we propose is new but builds directly on recent advances in panel unit root testing, and affords several advantages over some of the current panel unit root tests. Specifically, the SURADF test allows testing for a unit root in the real exchange rate of each country in the panel while at the same time accommodating contemporaneous cross-correlations of the regression residuals, heterogeneity in lag structures across the panel members, and different values of autoregressive coefficients among the panel members. Since the tests are conducted within a panel, the power advantages of hypothesis testing with T X N observations are exploited. One drawback of panel tests, including this one, is that unique results for a specific currency may not necessarily be generated across panels. This is because conclusions from the panel tests depend on panel size and composition, which are up to the discretion of the researcher. This problem is endemic to all other panel unit root tests.

In terms of power, simulations demonstrate that tests for a unit root conducted in a SUR panel framework are superior to independently-estimated, single-equation augmented Dickey-Fuller tests, with power gains up to a factor of seven for the panel of real exchange rates we considered. In addition, the SURADF test affords the possibility of distinguishing data generating processes across panel members whereas the MADF test, and all other panel unit root tests that are based on a joint null hypothesis, produce only oblique conclusions like "not all members of the panel contain a unit root" when rejection materializes.

The SURADF test is applied to three panels of real exchange rates -- the four-country group examined by Taylor and Sarno (1998), the EMS-7, and a group of twenty industrial countries. Applying the SURADF test to the four-country panel, three of the four real exchange rates show evidence of stationarity. Unfortunately, little support for purchasing power parity for dollar-based real exchange rates over the floating rate period is uncovered in the larger panels consisting of seven and twenty industrial countries. Furthermore, even with the significant improvement in power generated with the systems estimation of SUR, our results show broad agreement with conclusions from the pre-panel single-equation studies that use the less powerful ADF test. Our results are far less supportive of purchasing power parity than recent results that have emerged from those panel unit root studies that are based on tests of a joint hypothesis that all real exchange rates contain a unit root.

The SURADF procedure we propose and the results that are born from it contribute to the ongoing "search for stationarity" to pen Papell's (1997) title. The evidence from our test indicates that it may still be too early to claim widespread empirical support for Cassel's doctrine of purchasing power parity for the period of floating exchange rates.

References

Abuaf, Niso and Phillipe Jorion. (1990). "Purchasing Power Parity in the Long Run." *Journal of Finance*, 45, 157-74.

Breuer, Janice Boucher. (1994). "An Assessment of the Evidence on Purchasing Power Parity," in Estimating Equilibrium Exchange Rates, ed. John Williamson. Institute for International Economics. Washington, DC.

Breuer, Janice Boucher, Robert McNown, and Myles Wallace. (1998). "Misleading Inferences from Panel Unit Root Tests with an Illustration from Purchasing Power Parity." Department of Economics, University of South Carolina. mimeo.

Breuting, Jorg and Wolfgang Meyer (1991). "Testing for Unit Roots in Panel Data: Are Wages on Different Bargaining Levels Cointegrated?" Institut fur Quantitiatve Wirtschaftsforschung Working Paper.

Dickey, David. A. and Wayne Fuller (1979). "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Society*, 74, 427-31.

Frankel, Jeffrey and Andrew Rose. (1996). "A Panel Project on Purchasing Power Parity: Mean Reversion within and between Countries." *Journal of International Economics*. 40, 209-24.

Froot, Kenneth and Kenneth Rogoff. (1995). "Perspectives on PPP and Long-run Real Exchange Rates," in Handbook of International Economics, Volume III. eds. Gene Grossman and Kenneth Roggof. North-Holland. Amsterdam.

Greene, William H. (1997). "*Econometric Analysis*. Prentice Hall. Upper Saddle River, New Jersey.

Hakkio, Craig. (1984). "A Re-examination of Purchasing Power Parity." *Journal of International Economics*. 17. 265-77.

Im, Kyung So, M. Hashem.Pesaran, and Yongcheol Shin (1997). "Testing for Unit Roots in Heterogeneous Panels." Department of Applied Economics, University of Cambridge. mimeo.

Levin, Andrew and Chien-fu Lin. (1992). "Unit Root Tests in Panel Data: Asymptotic and Finite-sample Properties." University of California-San Diego Discussion Paper 92-23.

____________________________ (1993). "Unit Root Tests in Panel Data: New Results." University of California-San Diego Discussion Paper 93-56.

Maddala, G.S. and Shaowen Wu (1997). "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test." Ohio State University. mimeo.

McNown, Robert, and Myles Wallace (1989). "National Price Levels, Purchasing Power Parity, and Cointegration: A Test of Four High Inflation Economies," *Journal of International Money and Finance*. 8, 533-546.

O'Connell, Paul G. J. (1998). "The Overvaluation of Purchasing Power Parity." *Journal of International Economics*. 44, 1-19.

Oh, Keun-yeob (1996). "Purchasing Power Parity and Unit Root Tests using Panel Data." *Journal of International Money and Finance*. 15, 405-18.

Papell, David. (1997). "Searching for Stationarity: Purchasing Power Parity under the Current Float." *Journal of International Economics*. 43, 313-32.

Pedroni,Peter. (1995). "Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis." Indiana University Working Papers in Economics. 95-013.

Quah, Danny. (1990). "International Patterns of Growth: Persistence in Cross-country Disparities," MIT Working Paper.

Sarno, Lucio and Mark Taylor. (1998). "Real Exchange Rates Under the Recent Float: Unequivocal Evidence of Mean Reversion." *Economics Letters.* 60, 131-37.

Taylor, Mark. (1995). "The Economics of Exchange Rates." *Journal of Economic Literature*. 303,13-47.

Taylor, Mark and Lucio Sarno. (1998). "The Behavior of Real Exchange Rates during the Post Bretton Woods period." *Journal of International Economics*. 46, 281-312.

West, Kenneth D. (1988). "Asymptotic Normality when Regressors Have a Unit Root," *Econometrica*, 56, 1397-1418.

Wu, Yangru. (1996). "Are Real Exchange Rates Nonstationary? Evidence from Panel Data Tests." *Journal of Money, Credit, and Banking*. 28,54-63.

Wu, Shaowen and Jyh-Lin Wu. (1998). "Purchasing Power Parity under the Current Float: New Evidence from Panel Data Unit Root Tests." SUNY at Buffalo. mimeo.

Endnotes

1. Papell (1997) does not reject unit root behavior in a panel setting using quarterly data but is able to reject unit root behavior using monthly data.

2. The discussion that follows applies equally well to panel cointegration tests where the cointegrating regression residual is the variable whose time series properties are being investigated. For an example of panel cointegration tests, see Pedroni (1995).

3. Hakkio (1984) was the first to apply panel estimation in tests of purchasing power parity but he did not use unit root tests.

4. Levin and Lin's (1993) paper introduces treatment of serial correlation.

5. Although, in principle, the IPS test allows the lag structure to differ across panel members, the critical values they compute impose the same lag length across each panel member. They also find that overparameterizing the lag structure delivers a more favorable power/size tradeoff in their test than the LL test.

6. A nonzero drift term would impart a deterministic trend to the real exchange rate data, which is inconsistent with empirical evidence for most exchange rates. In addition, the ADF test statistic, and its counterpart estimated by SUR, is asymptotically normal if the intercept is included as the only deterministic component in both the DGP and the estimating equation (West, 1988).

7. The critical values for the MADF presented here differ from those generated by Taylor and Sarno (1998). Because their simulations did not set the intercepts to zero, their (implicit) null hypothesis for the four country panel was a unit root with drift, and their critical values were generated based on this assumption. The inclusion of the drift term appears inconsistent with most real exchange rate data, and also produces test statistics with asymptotic distributions related to the normal. Consequently, the critical values reported in Taylor and Sarno (1998) are close to corresponding chi-squared values and substantially smaller than the critical values reported here.

8. Both Finland and New Zealand were also significant at 0.10 in the individual ADF tests. In addition, a number of other countries in the twenty country panel have t-stats that are suggestive, but insignificant even at a generous 0.10 level.

9. The same conclusion holds when the tests are applied to panels of real exchange rates defined with respect to the German mark. The individual ADF tests reject the unit root hypothesis at the 10% level for only five of the twenty currencies. In the twenty country panel the SURADF test finds two countries to have stationary real exchange rates (Finland and France), and the French franc is the only stationary case in the other two panels. Further details on these German base currency results are available from the authors on request.

10. In panels consisting of ADF equations, the relation between error correlations and efficiency gain under SUR is complex. Since the ADF equation is a univariate autoregression, the correlation among error terms is related to the correlation among regressors. Below we prove that for a pair of AR1 processes with

identical autoregressive coefficients, the error correlations are identical to the correlations between the regressors. Since efficiency gains from SUR are weakened when the regressors from different equations are highly correlated, this offsets to some extent the efficiency gains due to high error correlations. Consider the two AR1 processes

$Y_{1,t} = a_1 Y_{1,t-1} + e_{1,t}$ and

$Y_{2,t} = a_1 Y_{2,t-1} + e_{2,t}$ where the AR coefficient, a_1, is the same for both series and $e_{1,t}$ and $e_{2,t}$ are white noise. Lagging one period and iterating backwards then

$Y_{1,t-1} = e_{1,t-1} + a_1 e_{1,t-2} + a_1^2 e_{1,t-3}$ .... and

$Y_{2,t-1} = e_{2,t-1} + a_1 e_{1,t-2} + a_1^2 e_{2,t-3}$ ....

The correlation coefficient is

$corr(Y_{1,t-1},Y_{2,t-1}) = (1 + a_1^2 + a_1^4 ..)cov(e_1,e_2)/\sqrt{}[(1 + a_1^2 + a_1^4 ..)\delta^2_{e1}]*\sqrt{}[(1 + a_1^2 + a_1^4 ..)\delta^2_{e2}]$

where δ^2_{e1} and δ^2_{e2} are the variances of e_1 and e_2. This reduces to

$cov(e_1,e_2)/\delta_{e1}\ \delta_{e2}$.

Note that this is identical to the correlation coefficient between $e_{1,t}$ and $e_{2,t}$. This result is valid even if $a_1 = 1.0$, and proves the proposition that for identical AR1 coefficients the cross correlation of the residuals is equal to the correlation of the regressors. Finally, when the series are of different orders of integration, I(1) and I(0), it is well known that asymptotically the correlation of the $Y_{1,t-1}$ and $Y_{2,t-1}$ will equal zero. It is obvious in this case that if the cross correlation of the residuals is not equal to zero it will not equal the regressor correlation, which is zero.

11. Note that the relationship between panel size and critical values holds for countries appearing in more than one panel. The critical values for Germany and France, which are the only two countries to appear in all three panels, increase as the panel size increases.

12. The relevant data for the regressions are found in Table 3. Critical values for countries included in more than one panel are entered each time they appear. So, for example, Germany which enters into all three panels is included in the regression three times. Each observation has a different critical value associated with a different panel size and average residual correlation.

13. There are exceptions to this tendency for the EMS seven-country panel, where for four of these countries the absolute critical values increase as the sample size increases from 250 to 500 observations.

14. Again, to minimize the sensitivity of the results to initial conditions, the first 50 observations on the dependent variable were omitted.

15. Restricting the intercepts to zero implies an equilibrium (mean) value of zero for stationary series. Since our real exchange rates are in logarithms, a long run equilibrium value of zero for the real exchange rate is equivalent to a value of 1.0 in the non-logged levels. Given the construction of real exchange rates using price indexes, imposing a long run mean of 1.0 is inappropriate.

16. The initial values of the I(1) series were set to zero while the I(0) series were initialized at their estimated mean values.

17. Jorion and Sweeney (1996) report a (panel) estimate of 0.97 for the G-10 countries. Papell reports single-equation estimates ranging between 0.84-0.97 with quarterly data for a group of twenty industrial countries, and Wu (1996) reports a (panel) estimate of 0.94 for quarterly data.

18. While not reported, but available on request, simulations indicate that the true critical values for the I(1) series in a mixed panel of I(1) and I(0) series depend primarily on the number of I(1) series, and not on the total number of members in the panel. Note that the rejection rates reported in Tables 8 and 9 for the I(1) series in the mixed panels are substantially below 0.05. The reason is that the true critical values in mixed panels are lower than those based on all series in the panel being I(1). However, because the researcher does not know the mix of the panel to start with the appropriate null is that all series are I(1), and we use critical values based on this null.

Table 1: Individual ADF Equation Results

Country Name	ADF	LAGS
1.Canada	-1.64	6
2. Australia	-2.01	3
3. Japan	-1.88	5
4. UK	-2.70c	7
5. Switzerland	-2.54	4
6. Germany	-2.56c	4
7. Austria	-2.39	4
8. Belgium	-2.41	4
9. Denmark	-2.05	3
10. Finland	-2.75c	7
11. France	-2.63c	4
12. Greece	-2.40	4
13. Italy	-2.49	4
14. Netherlands	-2.62c	4
15. Norway	-2.34	7
16. Portugal	-2.03	8
17. Spain	-2.63c	8
18. Sweden	-2.29	7
19. Ireland	-1.84	6
20. New Zealand	-2.74c	3

Notes: This table reports the individual country ADF statistics. LAGS = number of lagged augmentation terms.
c = significant at 0.10 level.

Table 2: ADF Tests under SUR Estimation and Critical Values				
SARNO & TAYLOR-4				
		CRITICAL VALUES		
Country Panel	SURADF	0.01	0.05	0.1
1. UK	-2.765	-4.194	-3.553	-3.200
2. Germany	-3.947b	-4.530	-3.946	-3.580
3. France	-4.057b	-4.560	-3.930	-3.584
4. Japan	-3.284c	-4.209	-3.581	-3.212
	MADF	CRITICAL VALUES		
		0.99	0.95	0.9
	23.673c	30.144	23.879	20.973
EMS-7				
1. Belgium	-4.274	-5.550	-4.880	-4.554
2. Denmark	-3.824	-5.417	-4.859	-4.506
3. France	-3.900	-5.555	-4.880	-4.555
4. Germany	-4.011	-5.550	-4.942	-4.642
5. Ireland	-2.406	-5.105	-4.480	-4.122
6. Italy	-2.907	-4.725	-4.069	-3.719
7. Netherlands	-4.039	-5.551	-4.959	-4.653
	MADF	CRITICAL VALUES		
		0.99	0.95	0.9
	23.062	43.883	36.403	32.985
20 INDUSTRIAL COUNTRIES				
1.Canada	-1.846	-4.637	-3.760	-3.370
2. Australia	-2.008	-4.671	-3.934	-3.534
3. Japan	-3.986	-5.137	-4.397	-4.001
4. UK	-2.695	-5.528	-4.726	-4.318
5. Switzerland	-4.676	-6.385	-5.562	-5.099
6. Germany	-5.973	-7.461	-6.777	-6.398
7. Austria	-5.805	-7.253	-6.554	-6.169
8. Belgium	-5.771	-7.129	-6.408	-6.006
9. Denmark	-4.995	-7.066	-6.302	-5.909
10. Finland	-5.249b	-5.789	-4.974	-4.495
11. France	-5.850	-7.277	-6.513	-6.082
12. Greece	-3.337	-6.181	-5.393	-4.938
13. Italy	-4.302	-5.794	-5.045	-4.626

14. Netherlands	**-5.869**	**-7.297**	**-6.593**	**-6.219**
15. Norway	**-4.717**	**-6.349**	**-5.579**	**-5.132**
16. Portugal	**-3.120**	**-5.843**	**-4.998**	**-4.564**
17. Spain	**-4.300**	**-5.886**	**-5.070**	**-4.614**
18. Sweden	**-3.593**	**-5.317**	**-4.489**	**-4.071**
19. Ireland	**-4.000**	**-6.564**	**-5.821**	**-5.402**
20. New Zealand	**-3.958c**	**-4.884**	**-4.164**	**-3.755**
	MADF	**CRITICAL VALUES**		
		0.99	**0.95**	**0.9**
	92.949	**119.718**	**105.969**	**98.687**

Notes: This table reports the results for the various SUR panels listed below. SURADF reports the estimated Dickey-Fuller statistics from the SUR regressions. The MADF statistics are listed separately. The three right hand side columns report the estimated critical values obtained from simulations based on a sample size of 91 and 10,000 replications.
a, b, and c denote significance at 0.01, 0.05, and 0.10 respectively.

Table 3: Average Residual Correlations

	Country	**Average Correlation**	**CV(.05)**
Panel A:	1. UK	0.625922	-3.553
	2. Germany	0.751879	-3.946
	3. France	0.761236	-3.929
	4. Japan	0.622313	-3.581
Panel B:	1. Belgium	0.910102	-4.899
	2. Denmark	0.905855	-4.858
	3. France	0.912930	-4.880
	4. Germany	0.913556	-4.942
	5. Ireland	0.866067	-4.480
	6. Italy	0.765170	-4.069
	7. Netherlands	0.921129	-4.959
Panel C:	1.Canada	0.061504	-3.760
	2. Australia	0.269665	-3.934
	3. Japan	0.532516	-4.396
	4. UK	0.569799	-4.726
	5. Switzerland	0.680782	-5.562

	6. Germany	0.734608	-6.777
	7. Austria	0.725770	-6.554
	8. Belgium	0.724767	-6.408
	9. Denmark	0.730742	-6.302
	10. Finland	0.622587	-4.974
	11. France	0.735409	-6.513
	12. Greece	0.669730	-5.393
	13. Italy	0.626953	-5.045
	14. Netherlands	0.731429	-6.593
	15. Norway	0.700363	-5.579
	16. Portugal	0.642511	-4.998
	17. Spain	0.643467	-5.070
	18. Sweden	0.533940	-4.488
	19. Ireland	0.709056	-5.821
	20. New Zealand	0.407895	-4.164

Notes: The average correlation between own country residuals from the SUR regressions with 91 observations and the other countries from the same panel are found under the column Average Correlation. The next column reports the 0.05 critical values for 91 observations.

Table 4: SUR ADF .05 CRITICAL VALUES/MADF .95 Critical Values

Sarno and Taylor-4				
	T=25	T=100	T=250	T=500
1. UK	-4.047	-3.531	-3.476	-3.428
2. Germany	-4.099	-3.885	-3.884	-3.860
3. France	-4.109	-3.892	-3.946	-3.907
4. Japan	-3.940	-3.515	-3.461	-3.423
MADF	31.992	23.543	23.008	22.387
EMS-7				
	T=25	T=100	T=250	T=500
1. Belgium	-5.191	-4.917	-4.903	-4.906
2. Denmark	-5.015	-4.786	-4.805	-4.831
3. France	-5.129	-4.872	-4.922	-4.915
4. Germany	-5.301	-4.913	-4.962	-4.942
5. Ireland	-4.867	-4.463	-4.473	-4.485
6. Italy	-4.453	-4.084	-4.019	-3.995
7. Netherlands	-5.257	-4.954	-4.962	-4.973
MADF	49.738	36.287	35.500	35.103
20 Industrial Countries				
	T=25	T=100	T=250	T=500

1.Canada	-5.937	-3.760	-3.414	-3.335
2. Australia	-6.028	-3.910	-3.588	-3.480
3. Japan	-6.456	-4.385	-4.096	-4.011
4. UK	-6.886	-4.692	-4.512	-4.419
5. Switzerland	-7.313	-5.493	-5.377	-5.280
6. Germany	-9.283	-6.711	-6.563	-6.516
7. Austria	-9.010	-6.493	-6.366	-6.314
8. Belgium	-8.795	-6.329	-6.234	-6.193
9. Denmark	-8.215	-6.215	-6.191	-6.519
10. Finland	-7.046	-4.877	-4.785	-4.719
11. France	-8.451	-6.466	-6.395	-6.342
12. Greece	-6.970	-5.298	-5.161	-5.108
13. Italy	-6.808	-5.006	-4.770	-4.708
14. Netherlands	-9.067	-6.555	-6.428	-6.388
15. Norway	-7.520	-5.559	-5.533	-5.532
16. Portugal	-7.304	-4.978	-4.840	-4.788
17. Spain	-7.308	-5.006	-4.938	-4.882
18. Sweden	-6.689	-4.453	-4.221	-4.159
19. Ireland	-7.732	-5.788	-5.751	-5.718
20. New Zealand	-6.183	-4.156	-3.850	-3.772
MADF	347.158	103.435	90.881	87.169

Notes: The SURADF and MADF critical values for various sample sizes (T) were obtained from simulations using 10,000 replications.

Table 5: Power Functions for the four country panel: 5,000 replications. All series I(0).

Rejection Rates (Power), T=25						
	ρ=0.99		**ρ=0.95**		**ρ=0.90**	
Country	**ADF**	**SURADF**	**ADF**	**SURADF**	**ADF**	**SURADF**
United Kingdom	5.1	5.6	6.1	7.6	7.1	8.1
Germany	4.4	5.1	5.3	5.8	5.8	8.4
France	4.4	5.1	5.4	6.4	6.4	8.8
Japan	5.1	5.4	5.6	7.0	7.0	8.4
MADF*	5.4		6.8		8.7	

Rejection Rates (Power), T=100						
	ρ=0.99		**ρ=0.95**		**ρ=0.90**	
Country	**ADF**	**SURADF**	**ADF**	**SURADF**	**ADF**	**SURADF**
United Kingdom	5.9	8.0	18.2	29.1	39.9	54.1
Germany	6.1	7.1	15.5	30.1	37.7	70.8
France	6.1	8.9	17.9	37.4	41.9	78.3
Japan	5.8	6.5	14.7	20.1	32.8	46.2
MADF*	8.6		45.1		89.6	

Rejection Rates (Power), T=250						
	ρ=0.99		ρ=0.95		ρ=0.90	
Country	ADF	SURADF	ADF	SURADF	ADF	SURADF
United Kingdom	11.5	17.3	74.3	87.0	96.9	99.0
Germany	8.6	13.7	63.5	93.3	97.6	99.9
France	9.1	14.9	69.6	95.0	98.4	99.9
Japan	7.9	10.7	59.2	75.0	96.4	98.6
MADF*	18.7		99.9		100.0	

Rejection Rates (Power), T=500						
	ρ=0.99		ρ=0.95		ρ=0.90	
Country	ADF	SURADF	ADF	SURADF	ADF	SURADF
United Kingdom	26.7	44.0	99.7	99.9	100.0	100.0
Germany	17.3	38.3	99.5	100.0	100.0	100.0
France	19.5	45.3	99.6	100.0	100.0	100.0
Japan	17.6	24.9	99.1	99.7	100.0	100.0
MADF*	61.2		100.0		100.0	

Notes: Rejection rates of the unit root null hypothesis are based on the simulated I(0) series having lag and covariance structures of the United Kingdom, Germany, France, and Japan. ρ is the value of the autoregressive coefficient on the simulated I(0) series.

*The null hypothesis for the MADF test is that *all* series are I(1). The power results are not comparable to the power results reported for the ADF and SURADF tests since the latter are tests on the individual series.

Table 6: Power Functions for the seven country EMS panel: 5,000 replications. All series I(0).

Rejection Rates (Power), T=25						
	ρ=0.99		ρ=0.95		ρ=0.90	
Country	ADF	SURADF	ADF	SURADF	ADF	SURADF
Belgium	4.6	5.6	5.0	8.5	5.9	11.9
Denmark	4.7	5.8	5.7	7.4	6.6	11.1
France	4.4	5.9	4.6	7.7	5.6	10.7
Germany	4.0	4.9	5.0	7.1	5.5	11.2
Ireland	3.5	4.7	3.7	4.9	4.2	6.4
Italy	3.9	5.7	4.7	7.7	4.9	9.1
Netherlands	4.3	5.9	4.7	7.7	5.7	11.2
MADF*	5.4		6.1		8.8	

Rejection Rates (Power), T=100						
	ρ=0.99		ρ=0.95		ρ=0.90	
Country	ADF	SURADF	ADF	SURADF	ADF	SURADF
Belgium	5.4	8.8	12.2	45.3	29.6	89.3
Denmark	5.5	9.2	13.1	40.9	32.0	84.4
France	5.7	9.8	12.8	43.2	27.8	85.2
Germany	5.7	9.4	11.9	44.4	26.7	89.0
Ireland	5.1	6.4	8.1	16.1	16.2	39.4
Italy	5.4	7.2	11.6	20.1	26.1	45.7

Netherlands	5.6	9.2	11.7	44.1	26.7	88.6
MADF*	8.7		44.1		92.3	
Rejection Rates (Power), T=250						
	ρ=0.99		ρ=0.95		ρ=0.90	
Country	**ADF**	**SURADF**	**ADF**	**SURADF**	**ADF**	**SURADF**
Belgium	7.6	20.6	50.9	98.9	94.1	100.0
Denmark	7.2	18.2	49.8	97.5	95.2	100.0
France	6.8	18.4	48.9	97.7	93.2	100.0
Germany	6.7	18.7	46.3	99.1	92.1	100.0
Ireland	6.1	9.2	29.0	67.0	83.2	98.0
Italy	6.9	10.9	45.0	72.3	91.3	98.0
Netherlands	7.2	19.2	46.3	99.1	92.2	100
MADF*	17.7		99.1		100.0	
Rejection Rates (Power), T=500						
	ρ=0.99		ρ=0.95		ρ=0.90	
Country	**ADF**	**SURADF**	**ADF**	**SURADF**	**ADF**	**SURADF**
Belgium	14.3	53.5	97.6	100.0	100.0	100.0
Denmark	13.7	46.2	97.4	100.0	100.0	100.0
France	13.7	49.3	96.8	100.0	100.0	100.0
Germany	12.8	50.5	96.5	100.0	100.0	100.0
Ireland	10.1	18.2	81.6	99.1	99.9	100.0
Italy	13.5	21.8	95.9	99.5	100.0	100.0
Netherlands	12.9	50.3	96.8	100.0	100.0	100.0
MADF*	54.1		100.0		100.0	

Notes: Rejection rates of the unit root null hypothesis are based on the simulated I(0) series having lag and covariance structures of the EMS-7. ρ is the value of the autoregressive coefficient on the simulated I(0) series.
*The null hypothesis for the MADF test is that *all* series are I(1). The power results are not comparable to the power results reported for the ADF and SURADF tests since the latter are tests on the individual series.

Table 7: Power Functions for the Panel of 20 Countries. All series are I(0). 5,000 replications.

Rejection Rates (Power), T = 25						
	ρ=0.99		**ρ=0.95**		**ρ=0.90**	
Range	**ADF**	**SURADF**	**ADF**	**SURADF**	**ADF**	**SURADF**
Lowest	3.9	4.7	4.2	5.5	3.8	6.5
Median	4.5	5.5	5.0	6.5	5.8	8.6
Highest	5.7	6.0	6.1	7.1	7.2	10.0
MADF*	5.9		7.0		8.3	
Rejection Rates (Power), T = 100						
	ρ=0.99		**ρ=0.95**		**ρ=0.90**	
Range	**ADF**	**SURADF**	**ADF**	**SURADF**	**ADF**	**SURADF**
Lowest	4.9	5.9	8.26	15.2	16.4	31.3

Median	5.7	8.0	12.3	24.8	27.8	50.6
Highest	6.5	11.9	20.3	48.1	40.9	93.4
MADF*	11.4		81.0		99.9	
Rejection Rates (Power), T = 250						
	ρ=0.99		**ρ=0.95**		**ρ=0.90**	
Range	**ADF**	**SURADF**	**ADF**	**SURADF**	**ADF**	**SURADF**
Lowest	6.5	5.7	29.9	55.1	72.9	91.1
Median	7.4	13.0	49.5	79.5	92.8	99.0
Highest	11.8	24.3	78.8	99.6	98.1	100.0
MADF*	45.2		100.0		100.0	
Rejection Rates (Power), T = 500						
	ρ=0.99		**ρ=0.95**		**ρ=0.90**	
Range	**ADF**	**SURADF**	**ADF**	**SURADF**	**ADF**	**SURADF**
Lowest	9.9	12.9	83.8	97.5	99.9	99.9
Median	13.6	29.4	97.3	99.9	100.0	100.0
Highest	32.3	57.7	99.9	100.0	100.0	100.0
MADF*	98.1		100.0		100.0	

Notes: The lowest, median, and highest rejection rates of the unit root null hypothesis are reported only. Rejection rates are based on the simulated series having lag and covariance structure of the twenty industrial countries. T indicates the number of time series observations in each regression. ρ is the value of the autoegressive coefficient for the simulated I(0) series.

*The null hypothesis for the MADF test is that *all* series are I(1). The power results are based on the critical value for all twenty series assumed I(0). The power results are not comparable to the power results reported for the ADF and SURADF tests since the latter are tests on the individual series.

Table 8: Power and Size Functions for Mixed Panel of 20 Countries with 4 I(0) and 16 I(1) processes. 100 observations. 5,000 replications.

Rejection Rates (Power) for the 4 I(0) cases						
	ρ=0.99		**ρ=0.95**		**ρ=0.90**	
Country	**ADF**	**SURADF**	**ADF**	**SURADF**	**ADF**	**SURADF**
United Kingdom	5.8	5.4	12.6	22.1	26.2	50.2
Germany	5.5	7.4	12.0	84.6	27.1	98.0
France	5.5	5.2	13.4	52.3	29.3	90.1
Japan	6.1	3.9	13.5	13.1	30.5	35.4
MADF*	10.8		81.9		99.0	

Low, Median, and High Rejection Rates (Size) across the 16 I(1) cases						
	ρ=0.99		ρ=0.95		ρ=0.90	
Range	**ADF**	**SURADF**	**ADF**	**SURADF**	**ADF**	**SURADF**
Lowest	4.6	0.7	4.6	0.1	4.6	0.1
Median	5.1	1.3	5.2	0.3	5.1	0.3
Highest	5.7	4.7	5.7	3.6	5.6	3.2

Notes: Rejection rates of the unit root null hypothesis are based on the simulated I(0) series having lag and covariance structures of the Sarno and Taylor-4 and the simulated I(1) series having lag and covariance structure of the remaining industrial 16. ρ is the value of the autoregressive coefficient on the simulated I(0) series. For the industrial 16 countries, we only report the lowest, median, and highest rejection rates across the sixteen countries.

*The null hypothesis for the MADF test is that *all* series are I(1). The power results are based on the critical value for all twenty series assumed I(0). The power results are not comparable to the power results reported for the ADF and SURADF tests since the latter are tests on the individual series.

Table 9: Power and Size Functions for Mixed Panel of 20 Countries with 7 I(0) and 13 I(1) processes. 100 observations. 5,000 replications.

Rejection Rates (Power) for the 7 I(0) cases						
	ρ=0.99		ρ=0.95		ρ=0.90	
Country	**ADF**	**SURADF**	**ADF**	**SURADF**	**ADF**	**SURADF**
Belgium	5.6	6.6	12.2	58.9	28.5	96.0
Denmark	6.2	5.8	12.0	41.8	8.8	87.8
France	5.7	6.2	13.0	49.9	26.0	90.8
Germany	5.3	7.2	11.8	84.3	17.9	99.6
Ireland	5.0	4.8	8.7	15.4	16.3	43.7
Italy	5.7	5.1	11.9	18.2	27.9	48.9
Netherlands	5.6	6.2	11.8	69.2	26.9	98.9
MADF*	8.9		79.6		99.6	

Low, Median, and High Rejection Rates (Size) across the 13 I(1) cases			
	ρ=0.99	ρ=0.95	ρ=0.90

Range	ADF	SURADF	ADF	SURADF	ADF	SURADF
Lowest	4.6	1.0	4.0	0.1	4.2	0.0
Median	4.9	2.4	4.8	1.0	4.6	0.6
Highest	5.7	5.1	5.4	3.6	5.2	4.1

Notes: Rejection rates of the unit root null hypothesis are based on the simulated I(0) series having lag and covariance structures of the EMS-7 and the simulated I(1) series having lag and covariance structure of the remaining industrial 13. ρ is the value of the autoregressive coefficient on the simulated I(0) series. For the industrial 13 countries, we only report the lowest, median, and highest rejection rates across the thirteen countries.

*The null hypothesis for the MADF test is that *all* series are I(1). The power results are based on the critical value for all twenty series assumed I(0). The power results are not comparable to the power results reported for the ADF and SURADF tests since the latter are tests on the individual series.

Table 10: Power and Size Functions for Mixed Panel of 20 Countries with 13 I(0) and 7 I(1) processes. 100 observations. 5,000 replications.

Low, Median, and High Rejection Rates (Power) for the 13 I(0) cases						
	ρ=0.99		ρ=0.95		ρ=0.90	
	ADF	SURADF	ADF	SURADF	ADF	SURADF
Lowest	5.1	2.8	10.7	10.3	19.4	26.7
Median	5.5	5.4	13.5	17.7	29.2	45.4
Highest	6.6	9.1	21.2	38.0	40.2	98.2
MADF*	12.2		95.9		100.0	

Low, Median, and High Rejection Rates (Size) across the 7 I(1) cases						
	ρ=0.99		ρ=0.95		ρ=0.90	
Range	ADF	SURADF	ADF	SURADF	ADF	SURADF
Lowest	4.3	0.7	4.0	0.0	4.4	0.0
Median	4.5	1.0	4.7	1.6	4.7	0.1
Highest	4.8	1.4	4.9	3.2	4.9	0.3

Notes: Rejection rates of the unit root null hypothesis are based on the simulated I(0) series having lag and covariance structures of the non EMS-7 and the simulated I(1) series having lag and covariance structure of the EMS-7. ρ is the value of the autoregressive coefficient on the simulated I(0) series. For the industrial 13 countries and EMS-7, we only report the lowest, median, and highest rejection rates across the panels.

*The null hypothesis for the MADF test is that *all* series are I(1). The power results are based on the critical value for all twenty series assumed I(0). The power results are not comparable to the power results reported for the ADF and SURADF tests since the latter tests are for the null hypothesis that a single member of the panel is I(1).