Forthcoming , *Journal of International Economics*

The Great Appreciation, the Great Depreciation, and the Purchasing Power Parity Hypothesis

David H. Papell

Department of Economics
University of Houston
Houston, TX 77204-5882
(713) 743-3807
E-Mail: dpapell@uh.edu

Revised: February 2001

Although there has been much recent work on Purchasing Power Parity (PPP), neither univariate nor panel methods have produced strong rejections of unit roots in U.S. dollar real exchange rates for industrialized countries during the post-1973 period. We investigate the hypothesis that these non-rejections can be explained by one episode, the large appreciation and depreciation of the dollar in the 1980s, by developing unit root tests which account for this event and maintain long-run PPP. Using panel methods, we can strongly reject the unit root null for those countries that adhere to the typical pattern of the dollar's rise and fall.

I am grateful to Jushan Bai, Fabio Canova, Charles Engel, Lutz Kilian, Chris Murray, Michael Palumbo, John Rogers, two anonymous referees, and to participants in seminars at Michigan, Ohio State, South Carolina, Texas Econometrics Camp III, the 1999 AEA meetings, and the CEPR/ONB Conference on Real Exchange Rates: Recent Theories and Evidence, for helpful comments and discussions.

1. Introduction

Purchasing Power Parity (PPP) is one of the most enduring topics in international economics, and the question of whether PPP holds during the post-Bretton-Woods system of flexible nominal exchange rates has been extensively analyzed. While the failure of PPP to hold in the short run was obvious after the first few years of generalized floating, long-run PPP has been subject to a "mean reversion in economic thought" (Lothian and Taylor, 1997). In the mid-1970s, models such as Dornbusch (1976) routinely used PPP as a long-run equilibrium condition. By the mid-1980s, the widespread failure to reject unit roots in real exchange rates led authors such as Stockman (1990) to construct models where long-run PPP did not hold. By the mid-1990s, however, research on both long-horizon data and on panels of post-1973 real exchange rates has led to a renewed belief in the validity of long-run PPP.

All variants of PPP postulate that the real exchange rate reverts to a constant mean. Evidence of long run PPP can be provided by tests of a unit root in the real exchange rate. If the unit root null hypothesis can be rejected in favor of a level stationary alternative, then there is long-run mean reversion and, therefore, long-run PPP.[1] The starting point for research on PPP during the current float is the observation that, using conventional Augmented-Dickey-Fuller (ADF) tests on univariate real exchange rates for industrial countries, the unit root null is rarely rejected. While these findings were initially taken as evidence against PPP, it has become clear that they say more about the low power of unit root tests with short time spans of data than about PPP.[2]

In response to these problems, research on long-run PPP has progressed in two directions. First, univariate techniques have been applied to long-horizon real exchange rates spanning one to two centuries. This data, however, combines periods of fixed and floating nominal exchange regimes, and cannot answer the question of whether evidence of PPP would be found with the same time span of flexible rates.[3] Second, tests for unit roots in panel data, notably those of Levin

[1] Breuer (1994), Froot and Rogoff (1995), and Rogoff (1996) survey various concepts of PPP.
[2] Froot and Rogoff (1995) and Lothian and Taylor (1997) discuss the implications for finding evidence of PPP of the low power of unit root tests with long half-lives and less than a century of data.
[3] In addition, if long-term real exchange rates either, as in Engel (2000), are generated from the sum of a random walk and a very volatile transitory component or, as in Hegwood and Papell (1998), contain permanent structural changes, rejection of the unit root null does not necessarily provide evidence of PPP.

and Lin (1992) and Im, Pesaran, and Shin (1997) have been used to test for PPP among industrialized countries in the post-1973 period.[4]

Panel unit root tests have not produced strong evidence of PPP for quarterly post-1973 U.S. dollar based real exchange rates. Papell (1997), using data for 21 industrialized countries from 1973 to 1994, cannot reject the unit root null at the 10 percent level when serial correlation is taken into account in calculating lag lengths and computing critical values. This result is unchanged when the sample is extended through 1996 in Papell and Theodoridis (1998). O'Connell (1998a), emphasizing contemporaneous correlation, also reports non-rejections of the unit root null for post-1973 real exchange rates.[5]

The non-rejections are not caused by the low power of panel unit root tests. Levin and Lin (1992) and Bowman (1999) both report very high size adjusted power for panels of the size, time span, and half-lives of the post-1973 real exchange rates. These power results, however, depend crucially on the assumption of independence across individuals, and are not applicable if cross-sectional correlation is present. Dollar-based real exchange rates are highly contemporaneously correlated. Engel, Hendrickson, and Rogers (1997) and O'Connell (1998a) stress the importance of adequately controlling for contemporaneous correlation in panel unit root tests involving real exchange rates. O'Connell proposes a maximum likelihood procedure to account for contemporaneous correlation, but at the cost of severely restricting the degree of serial correlation.

The behavior of most, but not all, dollar-based real exchange rates has been dominated by one episode: the large nominal appreciation and depreciation of the dollar in the 1980s. While there is widespread agreement that the initial appreciation of the dollar was caused by the monetary/fiscal policy mix of the United States in the late 1970s and early 1980s, there has been no successful explanation of the magnitude of the dollar's appreciation based on economic fundamentals. Frankel and Froot (1990), for example, suggest that the dollar "overshot the overshooting equilibrium." In the absence of fundamentals-based explanations, the appreciation is often described as a bubble, with a very rapid depreciation after the bubble burst.

[4] Another direction for research has been to use unit root tests with more power, notably the DF-GLS test of Elliot, Rothenberg, and Stock (1996). Application of these tests to post-1973 real exchange rates by Cheung and Lai (2000), however, produces only weak additional rejections of the unit root null among industrialized countries.

[5] Jorion and Sweeney (1996) and Papell (1997) report rejections of the unit root null with the German mark as the numeraire currency. Pedroni (1997) finds evidence of panel cointegration between post-1973 nominal exchange rates and relative prices. Several recent papers, including Higgins and Zakrajsek (1999) and Wu and Wu (1999) report stronger rejections with dollar real exchange rates using data through (at least) 1997.

The rise and fall of the dollar is illustrated, in Figure 1, by depicting nominal and real German mark and Norwegian krone exchange rates. The real exchange rate follows the pattern of appreciation and depreciation of the nominal exchange rate at high frequencies. Over longer horizons, the possibility of PPP emerges. It appears that the real exchange rate fluctuates around approximately the same mean in 1988 - 1996 as in 1973 - 1980. These short and long-horizon patterns, which are consistent with sticky-price models, characterize most dollar-based exchange rates of European countries. The patterns, however, are not universal. Figure 1 also depicts nominal and real Japanese yen and Australian dollar exchange rates. While the real and nominal exchange rates move together at high frequencies, the 1980 - 1987 episode is not so dominant and it does not appear that the pre-1980 and post-1987 real exchange rates fluctuate around the same (or necessarily any) mean.

We propose univariate and panel unit root tests for purchasing power parity that account for the appreciation and depreciation of the dollar.[6] The intuition that motivates the tests is to treat the rise and fall of the dollar in the 1980s as being determined outside the data generating process. The specific test allows for three changes in the slope, but no breaks in the intercept, of the trend function for the real exchange rates. We posit that the nominal bubble caused the real dollar to rise as if it had a deterministic trend. Once the bubble burst, the real dollar fell, again as if it had a deterministic trend. In theory, the first break would pick up the start of the dollar's appreciation at the beginning of the 1980s, the second break would be caused by the switch from appreciation to depreciation in 1984-85, and the third break would depict the end of the depreciation in 1987. In practice, because the break dates are determined endogenously, the breaks are not constrained to be anywhere near these dates.

We restrict attention to slope changes because post-1973 dollar-based real exchange rates are drawn from a nominal flexible exchange rate regime. Unlike nominal fixed exchange rate regimes, where devaluations and revaluations, especially following failed attempts to defend currencies, can lead to large discrete changes, both real and nominal exchange rates under nominal floating appear to be better characterized by long swings of appreciation and depreciation (slope changes) than by discrete jumps (intercept changes). In addition, we impose "PPP restricted broken trend" constraints to ensure consistency with long-run purchasing power parity. There is no time trend, producing a constant mean preceding the first break, the coefficients on the dummy

[6] Lothian (1998) has proposed the rise and fall of the dollar as an explanation for the difficulty in finding evidence of PPP, but does not develop formal tests.

variables which depict the breaks are restricted to produce a constant mean following the third break, and the pre-first and post-third break means are constrained to be equal.

We investigate these issues in four stages. First, we want to see if our assumption of three changes in the slope is supported statistically. We use tests for multiple structural changes, recently developed by Bai (1999), to determine the number of breaks in the U.S. dollar based real exchange rates of 20 industrialized countries. Allowing between zero and five slope changes, the model with three breaks is chosen for most of the countries.

Second, we use univariate methods, subject to the restrictions described above, to test the unit root null against the PPP restricted broken trend alternative for the 20 real exchange rates. The unit root null can be rejected (at the 10 percent level) for only one country. Given the lack of power in univariate unit root tests (even in the absence of structural change) with 25 years of data, we were not surprised that these tests provide little-to-no evidence of PPP. The median break dates are 1980 (III), 1985 (I), and 1987(II/III). Fifteen of the 20 real exchange rates exhibit breaks within (on average) one year of the median, with a typical pattern of appreciation of the dollar starting in 1980, depreciation starting in late 1984 or early 1985, and the end of the depreciation in 1987. Five of the countries, however, Australia, Canada, Greece, Japan, and Portugal, are clearly exceptions to the pattern.

Third, we develop panel unit root tests in the presence of PPP restricted structural change. Since the dates of the breaks are constrained to be the same across the different countries' real exchange rates, it is important that the assumption of a common break date be tenable. We construct various panels as follows: Starting with the full panel (20 real exchange rates), we subtract countries one-by-one in decreasing order of the distance between their break dates and the median break dates. Thus, as the size of the panels becomes smaller, the "commonality" of the break dates increases.

For the panels with 17 to 20 countries, where the common break date assumption is clearly violated, the unit root null cannot be rejected at the 10 percent level. For the panels with 11 to 16 countries, where the assumption of a common break date can be justified (with at most one exception), we reject the unit root null hypothesis in favor of the PPP restricted broken trend alternative at the 1 percent level. This provides very strong evidence that unit roots in post-1973 real exchange rates can be rejected once the great appreciation and depreciation of the dollar in the 1980s is taken into account. It should be emphasized, however, that the evidence applies only to those countries that adhere to the pattern of the dollar's rise and fall.

Estimating models which do not incorporate structural change provides additional evidence that the rise and fall of the dollar is the cause of previous non-rejections of the unit root hypothesis. Using conventional panel unit root tests with an alternative of level stationarity, we cannot reject the null for any of the panels. For the panels with 11 to 15 “typical” countries, as well as the panel of 16 countries, the evidence against unit roots is much stronger for the models with structural change. For the panels with 17 to 20 countries, which include more “atypical” countries, incorporating structural change does not strengthen the evidence against unit roots.

Fourth, we conduct Monte Carlo simulations to investigate the size and power of our tests. We perform simulations where all countries are stationary with PPP restrictions, all countries have unit roots with structural change, and with various combinations of stationary and nonstationary countries. The results in the paper are most consistent with the hypothesis that the "typical" countries, almost all European, are stationary while the "atypical" countries, mostly non-European, are nonstationary. They are clearly not consistent with either the hypothesis that PPP holds for all countries or the hypothesis that PPP does *not* hold for all countries. Furthermore, the results cannot be explained by the inclusion of a minority of stationary countries in the panels.

2. Univariate tests

The purpose of this paper is to investigate the purchasing power parity hypothesis among industrialized countries in the post-Bretton-Woods flexible exchange rate period. We use quarterly, nominal, end-of-period exchange rates and Consumer Price Indexes for industrialized countries, obtained from the International Monetary Fund’s International Financial Statistics (CD-ROM for 9-97). The data start in the first quarter of 1973 and end in the fourth quarter of 1996, providing 96 quarterly observations. There are 23 countries that are considered industrialized by the IMF. We do not use data for Iceland because of the existence of gaps in its CPI and for Luxembourg because it has a currency union with Belgium. The 21 remaining countries provide 20 real exchange rates with the U.S. dollar as the numeraire currency.[7]

The real (dollar) exchange rate is calculated as follows,

$$q = e + p^* - p \quad , \qquad (1)$$

[7] We do not extend the data past 1996 because data for 1997 and beyond reflects the actions taken by members of the European Union in 1997 to satisfy the Maastricht criteria for joining the Euro. Several of these criteria, including nominal exchange rates that satisfy the narrow EMS bands, low inflation, and limits on government budget deficits, have implications for purchasing power parity.

where q is the logarithm of the real exchange rate, e is the logarithm of the nominal (dollar) exchange rate, p is the logarithm of the domestic CPI, and p^* is the logarithm of the U.S. CPI.

The most common test for PPP is the univariate ADF test, which regresses the first difference of a variable (in this case the logarithm of the real exchange rate) on a constant, its lagged level and k lagged first differences,

$$\Delta q_t = m + a q_{t-1} + \sum_{i=1}^{k} c_i \Delta q_{t-i} + e_t \,, \qquad (2)$$

A time trend is not included in equation (2) because such an inclusion would be theoretically inconsistent with long-run PPP. The null hypothesis of a unit root is rejected in favor of the alternative of level stationarity if α is significantly different from zero. We use the recursive t-statistic procedure proposed by Hall (1994) to select the value of k, with the maximum value of k equal to 8 and the ten percent value of the asymptotic normal distribution used to determine significance.[8] The unit root null can be rejected (at the 5 percent level) for only one (the United Kingdom) out of 20 real exchange rates.[9]

A general principle of unit root tests, emphasized by Campbell and Perron (1991), is that nonrejection of the unit root hypothesis may be due to misspecification of the deterministic components included as regressors. Perron (1989) modeled this by allowing for a one-time exogenously determined break in the intercept, time trend, or both of the trend function, and the methodology has been extended to allow the breaks to be determined endogenously. These methods, however, are not suitable for testing purchasing power parity because the alternative hypothesis, (broken) trend stationarity, is not consistent with PPP.

Tests for a unit root in non-trending data which allow one break in the intercept, as in Perron and Vogelsang (1992), are consistent with a weaker version of PPP, called "qualified" PPP by Dornbusch and Vogelsang (1991) and "quasi" PPP by Hegwood and Papell (1998). The two-break unit root tests of Lumsdaine and Papell (1997) could also be extended to non-trending data. These tests could be used to model abrupt changes in real exchange rates, caused by devaluations and revaluations, during periods where nominal exchange rates are fixed. Post-1973

[8] As discussed by Campbell and Perron (1991) and Ng and Perron (1995), this procedure has better size and power properties than alternative methods such as selecting k based on information criteria.
[9] This result is so well known that we do not report the details.

real exchange rates, however, are not obviously characterized by a one (or two) time change in the intercept and do not appear to be well modeled by such tests.

Structural change in real exchange rates during periods of nominal floating appears to be better characterized by changes in the slope, rather than in the intercept, of the trend function. Periods of appreciation are followed by periods of depreciation, but the exchange rates do not rise and fall abruptly. This is related to the concept of "long swings" in exchange rates described by Engel and Hamilton (1990). Their objective, however, is to describe the behavior of nominal exchange rates, which they assume to be nonstationary. Under this assumption, they use Markov switching methods to search for changes in regimes of appreciation or depreciation (log first differences). Since our objective is to describe the behavior of real exchange rates, which we do not want to characterize as stationary or nonstationary a *priori*, we cannot use their methods.

Since the existent tests for a unit root in the presence of structural change are not well suited for investigating purchasing power parity in post-1973 real exchange rates, we need to develop an appropriate test. We proceed in two stages. First, independent of considerations involving PPP, we investigate whether our conjecture of three slope changes can be supported statistically. Next, we test for a unit root in the presence of a restricted version of the selected form of structural change, where the PPP hypothesis is embodied in the restrictions.

2.1 Testing for structural change

We use a likelihood ratio test recently developed by Bai (1999) to investigate structural change. His method allows for multiple structural changes, trending data, and lagged dependent variables. While

Bai's tests can allow for changes in both the intercept and the slope, we only allow for slope changes.[10]

2.1.1 Testing for the number of breaks

We start by estimating the following regression:

[10] While, for the reasons described above, we do not believe that post-1973 real exchange rates are well represented by intercept changes, we also estimated versions of Bai's test that allowed for either only intercept or both intercept and slope changes. The evidence of structural change with either specification was weaker than with the specification that included only slope changes.

$$q_t = m + bt + \sum_{i=1}^{p} g_i DTi_t + \sum_{j=1}^{k} c_j q_{t-j} + e_t \ , \qquad (3)$$

where the breaks occur at times TBi and the slope dummy variables DTi_t = (t - TBi) if t > TBi , 0 otherwise, i = 1,…, p. For each value of n (number of breaks), the optimal break(s) are chosen by minimizing the sum of squared residuals (SSR), where the breaks are chosen globally. The test statistic is based on the difference between the minimum SSR for n breaks and the minimum SSR for n + i breaks. The null hypothesis of n breaks is rejected in favor of the alternative hypothesis of n + i breaks if the test statistic is greater than the critical value. Bai shows that, starting with 0 breaks and increasing n by 1, the test procedure is consistent.[11]

Our procedure for determining the optimal number of breaks is as follows: We first test the alternative of one break against the null of zero breaks. If one break is significant, test two against one. If not, test two against zero. If two breaks are significant (from either test), test three against two. If not, test three against one or three against two (depending on the results of the previous tests.) Continue in the same manner until the maximum number of allowed breaks (5) has been reached.[12]

Estimation of multiple, globally chosen, breaks is computationally intensive. It was not possible, using available computers, to globally estimate more than three breaks. Instead, we utilize the following procedure, using the four break model as an example. We first choose (globally) the three breaks that produced the best fit in equation (3). Fixing the three breaks, we choose the fourth break that produced the best fit. Then, we set (new) TB1 = (old) TB2, (new) TB2 = (old) TB3, and (new) TB3 = (old) TB4. Fixing the three (new) breaks, we choose the "new" TB4 with the best fit. Repeating the process 10 times, we report the breaks chosen by the last iteration.[13]

2.1.2 *Calculation of critical values for the structural change tests*

The critical values for the structural change test depend on the specification of the model under both the null and the alternative: trending or non-trending data, number and types of

[11] The series does not have to be stationary in order for the test to be consistent.
[12] It is not advisable to stop after finding the first insignificant break because, in finite samples, it is possible that, if there are actually (for example) two breaks, tests of the one break alternative against the zero break null may not reject the null. Vogelsang (1998) discusses nonmonotonic power with two breaks.

dummy variables, and the values of k and of the c's. We first estimate equation (3), without breaks, for all 20 real exchange rates, and choose the optimal k by the BIC (with a maximum k = 5). Using this criterion, k = 1 for 18 of the 20 countries. With k chosen to equal 1, we again estimate (3) without breaks, and calculate the average values (across countries) of μ, β, and c_1. Using Monte Carlo methods on equation (3) with k = 1 and 96 observations (the exact size of our sample), we calculate critical values (with 5000 replications) for choosing between n and n + i breaks, with a minimum of zero and a maximum of five breaks.[14] The null hypothesis is n breaks and the alternative hypothesis is n + i breaks, with dynamics under both hypotheses determined by the average μ, β, and c_1 across countries.

2.1.3 Results of the tests for structural change

The results of the tests, which are presented in Table 1, show that the strongest evidence (lowest p-values) of structural change is for three breaks. Of the 20 real exchange rates, 12 have significant breaks at the 5 percent (or higher) level. Of these 12, nine are with 3 breaks. Four more are significant at the 10 percent level and, of these 4, two are with 3 breaks. Among the four which are not significant (at 10 percent), the lowest p-values for two of these are with 3 breaks. Overall, the strongest evidence of structural change for 13 of the 20 exchange rates is for three breaks, with the others mixed.

2.2 Testing for a unit root in the presence of restricted structural change

We now focus attention on testing for unit roots and purchasing power parity. We extend Perron's changing growth model, which allows a one-time change in the slope, but not in the intercept, of the deterministic trend to allow for three changes in the slope and be consistent with PPP.

---

[13] We did some experimentation, and the break dates usually converged after three or four iterations. We also used this method to search for three breaks, and it approximated the global search procedure very well.
[14] Calculation of the critical values is even more computationally intensive than estimation of the breaks. We utilize the same approximation method, but with only the first break chosen globally. Even with this approximation, it took about 12 hours to compute the critical values on a Pentium Pro 400. We did some

2.2.1 Construction of tests for a unit root

As described by Perron (1989,1997), since the changes in slope are presumed to occur instantaneously, the model is of the additive outlier type and is estimated by a two-step procedure. First, the series is detrended using the following regression,

$$q_t = m + g_1 DT1_t + g_2 DT2_t + g_3 DT3_t + z_t \ , \qquad (4)$$

For consistency with PPP, there is a constant mean (no time trend) prior to the first break. The breaks occur at times TB1, TB2, and TB3, and the dummy variables DTi_t = (t - TBi) if t > TBi , 0 otherwise, i = 1,…, 3. Testing for a unit involves estimating the following regression,

$$\Delta z_t = a z_{t-1} + \sum_{i=1}^{k} c_i \Delta z_{t-i} + e_t \ , \qquad (5)$$

This unit root test becomes a test of purchasing power parity by the addition of two restrictions,

$$g_1 + g_2 + g_3 = 0 \ , \qquad (6)$$

which imposes a constant mean following the third break, as well as prior to the first break, and

$$g_1(TB3 - TB1) + g_2(TB3 - TB2) = 0 \ , \qquad (7)$$

which restricts the mean following the third break to equal the mean prior to the first break. The null hypothesis of a unit root *without structural change* is rejected in favor of the alternative hypothesis of level stationarity *with PPP restricted structural change* if α is significantly different from zero in equation (5)

There are two possible methods for endogenously selecting the break dates: by choosing the breaks which minimize the t-statistic on α in equation (5) and by choosing the breaks which minimize the
sum of squared residuals (or maximize the joint F-statistic on DT1, DT2, and DT3) in equation (4). We use the second method for two reasons. First, since we want to investigate the proposition that the nonrejection of unit roots in real exchange rates is caused by the rise and fall of the dollar, we want the trend to fit the data as closely as possible. Second, minimizing the sum of squared residuals in (4) is computationally much less burdensome, which becomes important for simulating the critical values.

experimentation with globally chosen breaks and fewer replications. The approximation does not appear to have much effect on the critical values. In all of the simulations, we generate 146 observations and discard the first 50.

2.2.2 *Calculation of critical values for the unit root* tests

Choosing three breaks endogenously forces some compromises in simulating critical values for the unit root tests. We first choose the break that produced the best fit for the following equation,

$$q_t = m + g_1 DT1_t + z_t \quad , \qquad (8)$$

without imposing the PPP restrictions. Fixing the first break, we choose the second break, again without imposing the PPP restrictions, which produced the best fit. Then, fixing the first two breaks, we choose the third break, this time with the PPP restrictions, that produced the best fit. Finally, again with the PPP restrictions, we set (new) TB1 = (old) TB2 and (new) TB2 = (old) TB3, chose the "new" TB3 with the best fit, and repeat the process 10 times. Once the break dates are chosen, the series are detrended and the unit root test statistics calculated as in Equations (4) and (5).[15]

We calculate critical values using Monte Carlo methods. First we generate a unit root series (without structural change) with 96 observations (the actual size of our sample) and fit autoregressive (AR) models to the first differences of the data, using the BIC to choose the optimal AR model. Then we use the optimal AR model in order to generate the errors for our data. We use the optimal AR model with iid $N(0,\sigma^2)$ innovations to construct a pseudo sample of size equal to our sample. The test statistic is the t-statistic on *a* in Equation (5). The critical values for the finite sample distributions are taken from the sorted vector of 5000 replicated statistics. The critical values are about 50 percent higher, in absolute value, than critical values of ADF tests for non-trending data without breaks, but are within the span of the various critical values for tests of a unit root in the presence of structural change in Perron and Vogelsang (1992) and Perron (1997).[16]

2.2.3 *Results of the univariate unit root tests*

---

[15] The model with 96 observations takes about 10 minutes to estimate on a Pentium Pro 400, which would require 5 weeks to compute 5000 replications. These calculations, in contrast, took less than one day. We computed 1000 replications for the model where the three breaks were chosen globally, and the critical values were very close to those that we report.

[16] While we did not want to calculate 20 sets of critical values (one for each series), we did some experimentation with computing critical values based on AR models of the first differences of the actual data. As in Papell (1997), this made little difference for post-1973 real exchange rates.

The results of the univariate unit root tests, reported in Table 2, provide little-to-no evidence of PPP. The unit root null cannot be rejected at the 5 percent level for any country, and can be rejected at the 10 percent level only for France. This should not be surprising. Univariate ADF tests have very low power in samples of this size and time span, and there is no reason to believe that univariate tests for unit roots in the presence of structural change would have greater power.[17]

The dates of the breaks and the coefficients on the dummy variables are also reported in Table 2. The median values of the breaks are 1980(III), 1985(I), and 1987(II/III). The typical pattern of breaks, reflecting the rise and fall of the dollar in the 1980s, is shared by most of the real exchange rates, with the coefficient γ_1 on the first break dummy variable positive (appreciation of the dollar), the coefficient γ_2 on the second negative (depreciation of the dollar), and the coefficient γ_3 on the third positive (end of the depreciation). For the countries which adhere to this pattern, γ_2 is generally at least twice as large (in absolute value) as γ_1, reflecting that the fall of the dollar was much faster than its rise.

In order to provide a measure of how well the individual real exchange rates fit the typical pattern, we calculate the root median squared error (RMSE) of the deviations between the break dates for each country and the median break dates, and report the results in Table 2. The RMSE is defined as the square root of the sum (over the three breaks) of the squared deviations (measured in quarters) from the median values of the breaks.[18] We classify the 15 countries with the smallest RMSE's as typical. These real exchange rates adhere to the pattern of the rise and fall of the dollar and have breaks that are, on average, less than one year away from the median. Japan is the most obvious exception to the typical pattern, with the highest RMSE and the opposite pattern of coefficients on all three dummy variables. The other atypical countries are (in order) Australia, Portugal, Canada, and Greece.

While the classification between typical and atypical countries is partly arbitrary, we can provide some justification. The difference between the largest RMSE for a typical country (Switzerland and the United Kingdom) and the smallest RMSE for an atypical country (Greece) is 6.48. This is much larger than either the difference, 1.49, between the RMSE's for Switzerland and the United Kingdom and the next largest RMSE for a typical country (Italy) or the difference,

[17] Since the chosen value of k was equal to the maximum (8) in several cases, we estimated the models with the maximum raised to 12. The results were unchanged.

[18] We use the median, rather than the mean, to mitigate the effect of outliers and the RMSE, rather than the sum of the absolute value of the deviations, to weight one large deviation more heavily than several smaller deviations.

0.88, between the RMSE for Greece and the next smallest RMSE for an atypical country (Canada).[19]

The univariate results for a selection of four typical countries: Germany and Norway, which have very small RMSE's, and Switzerland and the United Kingdom, which have the highest RMSE's among the typical countries, are depicted in Figure 2. The figures (not shown) for the other typical countries closely resemble those for Germany and Norway. The real exchange rates appreciate against the dollar in

the late 1970s below the pre-first-break mean, and then follow the cycle of appreciation and depreciation until the third break in 1987. The results for the four most atypical countries: Australia, Canada, Japan, and Portugal, are depicted in Figure 3. The breaks are often far removed from the typical pattern of the rise and fall of the dollar. In the case of Japan, the two breaks at the very end of the sample appears to reflect the consequences of imposing mean reversion on trending data.

[19] Canova (1997) proposes a method for determining the number of groups and the location of break points in the cross sectional dimension of a panel. His results, however, are for panels with a cross section dimension that is much larger, both in absolute terms and relative to the time series dimension, than our data.

3. Panel tests

The low power of unit root tests against highly persistent alternatives with anything less than a century of data has inspired the development of panel unit root tests which exploit cross section, as well as time series, variation. Variants of these tests have been developed by Levin and Lin (1992) (LL), Im, Pesaran, and Shin (1997) (IPS), Maddala and Wu (1999), and Bowman (1999). Applications of these tests to post-1973 real exchange rates of industrialized countries include Abuaf and Jorion (1990), Frankel and Rose (1996), Jorion and Sweeney (1996), Oh (1996), Wu (1996), O'Connell (1998a), Papell (1997), Papell and Theodoridis (1998,2000), and Wu and Wu (1999).[20]

A panel extension of the univariate ADF test in Equation (2), which accounts for both a heterogeneous intercept and serial correlation, would involve estimating the following equations,

$$\Delta q_{jt} = \mu_j + \alpha q_{jt-1} + \sum_{i=1}^{k} c_{ij} \Delta q_{jt-i} + \varepsilon_{jt} \quad , \qquad (9)$$

where the subscript j indexes the countries, and μ_j denotes the heterogeneous intercept. The test statistic is the t-statistic on α. The null hypothesis is that all of the series contain a unit root and the alternative hypothesis is that all of the series are stationary. In Papell (1997), we estimate Equation (9) using feasible GLS (seemingly unrelated regressions), with α equated across countries and the values for k taken from the results of univariate ADF tests.[21] For quarterly data with a panel of 20 industrialized countries, we could not reject the unit root null at the 10 percent level with the U.S. dollar as the numeraire currency, but could reject the null at the 1 percent level with the German mark as numeraire.[22]

The non-rejections of unit roots in post-1973 real exchange rates with the dollar as numeraire are not caused by low power of panel unit root tests. In Papell (1997), we estimate a

---

[20] This is an incomplete list, and does not include studies that use data from developing countries, tradable goods prices, panel cointegration, etc.

[21] Im, Peseran, and Shin (1997) and Maddala and Wu (1999) develop tests where α can vary across countries. The alternative hypothesis for these tests is that at least one of the series is stationary. Based on the results of univariate ADF tests in Papell (1997), this does not appear to be important for our sample of dollar-based real exchange rates of 20 industrialized countries. With univariate ADF tests, α is always negative but rarely significantly different from 0. If the α's cannot be shown to be significantly different from 0, it is difficult to see how a compelling case can be made that they are significantly different from each other. Bowman (1999) shows that size adjusted power falls much faster for the LL test than for the IPS test when only a subset of the members of a panel are stationary. Since rejection of the unit root null is normally interpreted as evidence that all real exchange rates are stationary, we view this as an advantage of the LL tests.

[22] Papell and Theodoridis (2000) estimate panels with all 21 industrialized countries as numeraire, and find strong rejections of unit roots in real exchange rates for most European countries.

value of α of -0.069 for a panel of 20 industrialized countries. Bowman (1999) reports that, for a panel of 20 members with 100 observations, the 5 percent size adjusted power of the LL test, with α equaling -0.05 for each member, is 0.99. The opposite results of panel unit root tests with the dollar and mark as numeraire are not caused by serial correlation. While Im, Peseran, and Shin (1997) report a loss of power with serial correlation, and Papell (1997), choosing the value of k by the data dependent methods described above, calculates critical values which are 15 percent larger (in absolute value) than those in Levin and Lin (1992), there is no difference in the amount of serial correlation between dollar and mark real exchange rates.

The clearest difference between dollar and mark real exchange rates involves contemporaneous, or cross-sectional, correlation. The large appreciation and depreciation of the dollar in the 1980s produces more commonality in dollar than in mark real exchange rates. In practice, there is a trade-off between accounting for serial and contemporaneous correlation. O’Connell (1998a) proposes a maximum likelihood estimator to account for contemporaneous correlation which, given the sample size of quarterly post-1973 data, severely restricts the degree of allowable serial correlation. Papell (1997) uses a feasible GLS (SUR) estimator which accounts for contemporaneous correlation and allows for more flexibility in modeling serial correlation, but does not iterate to maximum likelihood.[23]

3.1 Construction of panel unit root tests in the presence of restricted structural change

We extend the unit root tests in the presence of restricted structural change, developed above, to the panel context. The dates of the breaks are first chosen by using the following feasible GLS (SUR) regressions,

$$q_{jt} = m_j + g_1 DT1t + g_2 DT2_t + g_3 DT3_t + z_{jt} \quad , \qquad (10)$$

subject to the PPP restrictions described in Equations (6) and (7), where the dates of the breaks are chosen endogenously to maximize the joint log-likelihood. At this stage, while the intercepts are heterogeneous, the coefficients on the dummy variables are constrained to be equal across countries.[24]

[23] O’Connell (1998a) shows that, if there is no serial correlation or if both the lag lengths and the values of the c’s are the same for each country, panel unit root tests of real exchange rates using GLS are invariant to the choice of numeraire currency. These restrictions, however, are rejected by both O’Connell (1998b) and Papell and Theodoridis (2000).

[24] As in the univariate case, the model is of the additive outlier type and is estimated by a two-step procedure. The coefficients on the trend and the dummy variables are equated across countries at this stage to decrease

Once the break dates are chosen, the series are detrended as follows,

$$q_{jt} = m_j + g_{j1}DT1t + g_{j2}DT2_t + g_{j3}DT3_t + z_{jt} \quad , \tag{11}$$

where the coefficients on the dummy variables are now allowed to vary across countries. The test statistic is the t-statistic on α in the following feasible GLS (SUR) regressions,

$$\Delta z_{jt} = a z_{jt-1} + \sum_{i=1}^{k} c_{ij} \Delta z_{jt-i} + e_{jt} \ , \tag{12}$$

The null hypothesis is that all of the series have a unit root *without structural change* is rejected against the alternative hypothesis that all of the series are stationary *with PPP restricted structural change* if α is significantly different from zero. The dates of the breaks and value of α are constrained to be equal across countries, but the intercepts, coefficients on the dummy variables, and the values of the *k*'s and the *c*'s are heterogeneous. Purchasing power parity under the alternative is imposed by the restrictions,

$$g_{j1} + g_{j2} + g_{j3} = 0 \quad , \tag{13}$$

and,

$$g_{j1}(TB3 - TB1) + g_{j2}(TB3 - TB2) = 0 \quad , \tag{14}$$

for each country. This imposes a constant mean prior to the first break and following the third break, and constrains the means to be equal.

3.2 Calculation of critical values for the panel unit root tests

We calculate critical values for the panel unit root tests using Monte Carlo methods. For each of the panels, we fit univariate autoregressive (AR) models to the first differences of the real exchange rates, treat the optimal estimated AR models as the true data generating processes for the errors in each of the series, and construct real exchange rate innovations from the residuals.[25] We then calculate the covariance matrix Σ of the innovations. We use the optimal AR models with iid N(0,Σ) innovations to construct pseudo samples of size equal to the actual size of our series (96 observations). Since Σ is not diagonal, this preserves the cross-sectional dependence

---

computation time, which becomes necessary for calculating the critical values. We did some experimentation with allowing these coefficients to vary across countries, and the choice of breaks was not affected.

[25] We use the BIC to choose the optimal AR model. While it would be desirable to allow the first differences of the other real exchange rates to enter into the AR model, the size of the cross-section relative to the number of observations makes this infeasible.

found in the data. We then take partial sums so that the generated real exchange rates have a unit root without structural change by construction.

As in the univariate case, choosing three breaks endogenously in the panel context forces compromises in calculating critical values. We first choose the break that produced the best fit for the following set of equations,

$$q_{jt} = \mu_j + \gamma_1 DT1_t + z_{jt} \quad , \tag{15}$$

without imposing the PPP restriction. Following the procedure for the univariate case, we then choose the second and (imposing the PPP restriction) third breaks, and iterate. Once the break dates are chosen, the series are detrended and the unit root test statistics calculated as in Equations (11) and (12). The critical values for the finite sample distributions are calculated as described above for the univariate model with 5000 replications for the exact number of countries and time span of each panel.[26] Critical values with and without breaks are reported in Table 3.[27]

3.2 Results of the panel unit root tests

The results of the panel unit root tests in the presence of restricted structural change are described in Table 3. We estimate models for the full panel of 20 real exchange rates and, excluding countries one-by-one in decreasing order of their RMSE's, panels consisting of between 11 and 19 countries.[28] The smaller panels of between 11 and 16 countries, which include (at most) one atypical country, provide very strong evidence of purchasing power parity. The unit root null can be rejected at the 1 percent level, for each of the seven panels. The larger panels of between 17 and 20 real exchange rates, which include more of the atypical countries, provide no evidence of PPP. The unit root null cannot be rejected at standard significance levels for any of the panels, with all but one of the p-values above .20. The transition from the panel with 17 countries to the panel with 16 countries is particularly striking. The break dates become very

---

[26] The model with 15 countries and 96 observations takes about 40 minutes to estimate on a Pentium Pro 400, which would require 20 weeks to compute 5000 replications. These calculations took between one and two days. We computed 500 replications for the model with 15 countries where the three breaks were chosen globally, and the critical values were very close to those that we report.

[27] A potential problem with the critical values is that they do not incorporate our method of selecting the panels of real exchange rates out of the group of 20 countries. In order to correct for this, we generated 20 series of 96 observations, calculated three breaks for each series using the procedure described above, chose the 15 series with the smallest RMSE's, and calculated critical values for the panel with those 15 series. These critical values were very close to the critical values for the panel of 15 countries reported in Table 3.

[28] Since Switzerland and the United Kingdom have the same RMSE, we use root mean cubed error as the tiebreaker.

close to the median break dates from the univariate models, and the p-values for the unit root tests fall from .224 to .008.

The break dates and coefficients for the smaller panels, reported in Table 3, reflect the rise and fall of the dollar. The breaks are very tightly clustered. The first break occurs in 1980 (II) or (III), the second in 1985 (I), and the third in 1987 (III) or (IV) for all seven panels. The coefficients γ_1 and γ_3 are positive, while γ_2 is negative and more than twice as large as γ_1, for all of the panels. This is consistent with the pattern of sharp real appreciation of the dollar from 1980 to 1985, followed by even sharper real depreciation until 1987, that was observed in the univariate estimates.[29] The break dates and coefficients for the larger panels, which do not produce evidence of PPP, do not reflect the dollar's rise and fall.

How can we be sure that previous non-rejections of unit roots in real exchange rates, reported above, are caused by the rise and fall of the dollar in the 1980s? It is possible that our choice of countries is simply a fortuitous selection, which is particularly favorable to the PPP hypothesis. We investigate this by estimating a "standard" panel unit root model that, as described in Equation (10), does not account for structural change, for the same panels of 11 to 20 countries. The results are also reported in Table 3. The unit root null cannot be rejected at the 5 percent level for any of the panels, and can only be rejected at the 10 percent level for one panel. This is consistent with previous failures to find panel evidence of PPP with the U.S. dollar as the numeraire currency.

It is useful to examine the p-values as the size of the panels is reduced. In the absence of breaks, the p-values do not obviously rise or fall as the number of countries falls. Selecting panels of typical countries for which the breaks are closer to the median does not provide stronger evidence of purchasing power parity unless the effects of the rise and fall of the dollar are taken into account. For the panels between 17 and 20 countries, the p-values are all smaller without breaks than with the three restricted breaks. For the smaller panels between 11 and 16 countries, the opposite occurs. The p-values are all much smaller with the PPP restricted breaks.

Several recent papers, including Obstfeld and Taylor (1997) and Taylor and Peel (2000), have provided evidence of nonlinear mean reversion of real exchange rates during the post-Bretton-Woods period. These studies find that the speed of convergence to PPP increases with the distance of real exchange rates from their means. At first glance, our finding that the evidence of PPP strengthens once the largest deviation, the rise and fall of the dollar in the 1980s, is taken

[29] These coefficients are from the estimates in Equation (11), and are constrained to be equal across countries.

into account, appears to contradict these results. Examination of the coefficient γ_2 in Table 3, however, shows that, once the dollar peaks in 1985(I), there is very fast mean reversion for the panels of 11 to 16 countries for which the unit root null can be rejected. This evidence of fast mean reversion for the largest PPP deviation is consistent with the hypothesis of nonlinear mean reversion.

4. Size and power of the panel unit root tests with structural change

Using panel methods, we have obtained strong rejections of unit roots in real exchange rates in favor of a PPP restricted broken trend alternative for the panels with between 11 and 16 countries, but no rejections for the panels with between 17 and 20 countries. What do these findings mean? Do they provide evidence of PPP for all 20 countries, 16 of the 20, none of the 20, or some other subset? We proceed to investigate the size and power of our unit root tests in order to interpret the results.[30]

4.1 Construction of the size and power tests

We restrict attention to a particular class of data generating processes. For each country j, we construct the "real exchange rate" q as a combination of two processes:

$$q_{1jt} = \mu_{1j} + \gamma_{j1} DT1_t + \gamma_{j2} DT2_t + \gamma_{j3} DT3_t + \varepsilon_{1jt} \quad , \tag{16}$$

and,

$$q_{2jt} = \mu_{2j} + \alpha_j q_{2jt-1} + \varepsilon_{2jt} \quad . \tag{17}$$

If $\alpha < 1$ and the restrictions from (13) and (14) are imposed, the real exchange rate is stationary and PPP holds. If $\alpha = 1$ and the restrictions from (13) and (14) are not imposed, the real exchange rate has a unit root component and PPP does not hold.

The power of our unit root tests can be investigated by constructing panels with artificial data under the alternative hypothesis where all countries are stationary with PPP restricted structural change, performing our unit root tests on these constructed panels, and tabulating how often the unit root null is (correctly) rejected. While the tests are (by use of bootstrap critical values) correctly sized under the null hypothesis of a unit root *without* structural change, they are

[30] Breuer, McNown, and Wallace (2000) investigate size distortions in panel unit root tests that do not incorporate structural change with mixed panels.

not necessarily correctly sized under the null of a unit root *with* structural change.[31] We investigate potential size bias by constructing panels with artificial data under the hypothesis that all countries have a unit root with structural change, performing our unit root tests on these constructed panels, and tabulating how often the unit root null is (incorrectly) rejected.

We construct panels of $q_{jt} = q_{1jt} + q_{2jt}$ that simulate both stationary and unit root processes. For the stationary series, we generate q_{1jt} by imposing the DT's country-by-country from Table 2 and estimating the γ's subject to (13) and (14). This incorporates both structural change and the PPP restrictions. For the unit root series, the same DT's are imposed but the γ's are estimated unrestricted. This incorporates structural change but not the PPP restrictions. The q_{2jt} are generated with values of $\alpha = .97$ for the stationary series and with values of $\alpha = 1$ for the unit root series.[32] The covariance matrix for q_{1jt} is computed from the residuals of (4), with the PPP restrictions (6) and (7) imposed for the stationary, but not the unit root, series. The covariance matrix for q_{2jt} is computed from the innovations of the actual data for both types of processes.

4.2 Results for homogeneous panels

We report simulations where all of the countries are either stationary or contain a unit root in Table 4. The columns labeled "Stationary with PPP Restrictions" report the fraction (at various significance levels with 1000 replications) that the unit root null can be rejected against the PPP restricted alternative if the data is generated using (16) and (17) with $\alpha = .97$ and the PPP restrictions imposed. The tests have good power when all of the series are stationary. The 5% size adjusted power is over .80 and the 10% size adjusted power is over .90 for all of the panels. Even the 1% size adjusted power is over .70

[31] The tests are also not necessarily correctly sized if the data is generated as a sum of a stationary and a unit root process as in Engel (2000).
[32] We estimated an AR (1) for each of the 20 real exchange rates, and the average value of α was .937. The value of $\alpha = .97$ for the stationary series was chosen by the methods described in Andrews (1993) to correct for the downward bias in these estimates so that the generated and actual series would have the same persistence.

for all but one panel. The columns labeled "Unit Root with Structural Change" report the rejection fractions if the data is generated using (16) and (17) with α = 1.00 with no PPP restrictions. The tests are oversized when all of the series contain a unit root, with true sizes over .40 at 5% nominal size and over .50 at 10% nominal size.

The central empirical result of the paper is that the unit root null can be rejected in favor of a PPP restricted alternative at the 1% level for panels of 11 - 16 countries but cannot be rejected at the 10% level for panels of 17 - 20 countries. This result is not consistent with the hypothesis that the data for all of the countries is generated by either of these specifications. Suppose that all of the real exchange rates were stationary. This would be consistent with the rejections for the panels of 11 - 16 countries, where the 1% size adjusted power is between .786 and .815. It would not, however, be consistent with the failure to reject for the panels of 17 - 20 countries. For each of these panels, the 10% size adjusted power is over .90 yet the unit root null is not rejected. Now suppose that all of the countries contain a unit root with structural change (α = 1.00). This would not be consistent with the 1% rejections for the panels of 11 - 16 countries, since the true size with a nominal size of 1% is between .255 and .342. It would be consistent with the failure to reject for the panels of 17 - 20 countries, since the true size with a nominal size of 10% is between .491 and .524. It is impossible, however, for all members of the panels of 17 - 20 countries, but not all members of the panels of 11 - 16 countries, to contain a unit root when the latter are subsets of the former.

4.3 Results for mixed panels

We report the power of panel unit root tests with a mix of stationary and unit root countries in Table 5. For each panel of 11 - 20 countries, we perform simulations with at least eight stationary countries. As above, "stationary" denotes α = .97 with the PPP restrictions and "unit root" denotes α = 1.00 without the PPP restrictions. The power of the tests falls with the inclusion of even a few unit root countries. The 1% size adjusted power is below .50 with four or more unit root countries for about half of the panels, and below .50 with six or more unit root countries for all of the panels. As would be expected, the 10% size adjusted power is much higher, above .50 for all panels.

What does this tell us about our empirical results? We rejected the unit root null at the 1% level for all panels of 11 - 16 countries. We will call a rejection "consistent" with a mix of stationary and unit root countries if the 1% size adjusted power for that mix is above .50, and "inconsistent" if it is below .50. The rejections are generally consistent with panels containing 11 or more stationary countries, but inconsistent with panels containing 10 or fewer stationary countries. In particular, the rejections are not consistent with the hypothesis that the results are driven by the inclusion of a few stationary countries.

For the panels with eight stationary countries, the 1% size adjusted power is below .40 for all but one of the panels and below .50 for the other. We also failed to reject the unit root null for the panels with between 17 and 20 countries. Since the 10% size adjusted power for these panels with as few as eight stationary countries is above .50, this provides no additional information beyond the previous results.

We have shown that our empirical results are not consistent with either the hypothesis that all of the real exchange rates are stationary or with the hypotheses that they all contain a unit root. Furthermore, the findings with a mix of stationary and unit root series restrict the consistent results to between 11 and 16 stationary real exchange rates. We know that, while the tests have sufficient power, they are oversized when all of the countries contain a unit root. We would like for the tests to reject the unit root null when all of the series are stationary and to fail to reject the null when any of the series contain a unit root. By this criterion, they are also oversized with mixed panels. The true size of the tests with one unit root country at 1% nominal size is over .50 for all panels, and remains over .50 for most of the panels with as many as four unit root countries.

In order to develop a test with better size properties, we investigated what value of α would produce a correctly sized test when all countries were specified identically, and found that $\alpha = 1.014$ in (17) with no PPP restrictions in (16) provided the best approximation. We denote this specification as "nonstationary". We report the power of panel unit root tests with a mix of stationary and nonstationary countries in Table 6. These tests are much better sized. The true size of the tests with one nonstationary country at 1% nominal size is below .50 for all except one of the panels. At 5% nominal size, the true size is below .50 with two nonstationary countries for all panels. Even at 10% nominal size, the true size is below .50 with two nonstationary countries for all but three panels and is below .50 with three nonstationary countries for all panels.

We proceed to interpret our empirical results in the context of these findings. Recall that we rejected the unit root null at the 1% level for all panels of 11 - 16 countries. Since the 1% size adjusted power is between .218 and .478 for these panels, the results are inconsistent with the hypothesis that even one of these real exchange rates contains a unit root.[33] Furthermore, since the only examples of 1% size adjusted power above .40 are for the panels of 15 and 16 with one stationary country, the results are clearly inconsistent with fewer that 14 stationary countries. We also failed to reject the unit root null at the 10% level for the panels of 17 - 20 countries. Based on the 10% size adjusted power, these results are consistent with 17 or fewer stationary countries for the panels of 19 and 20, 16 or fewer stationary countries for the panel of 18, but only with 14 or fewer stationary countries for the panel of 17. In
particular, the 10% size adjusted power for the panel of 17 with one nonstationary country is .800, clearly inconsistent with the failure to reject the unit root null for that panel if there were truly 16 stationary countries. Combining the rejections and failures to reject, the empirical results are most consistent with the hypothesis that 15 of the 20 countries are stationary. They are clearly inconsistent with the either the hypothesis that more than 16 or fewer than 14 of the countries are stationary.

5. Conclusions

The proliferation of recent work on purchasing power parity underscores its importance as a central topic in international economics. The development of panel unit root tests presents both a challenge and an opportunity for researchers attempting to find strong evidence of long-run purchasing power parity using data from the current float. The opportunity occurs because, in contrast with univariate methods, panel unit root tests with 20 individuals and 100 quarterly observations have sufficient power to reject the unit root null in favor of a level stationary alternative, even with half-lives of over four years. The challenge arises because, again in contrast with univariate methods, failure to reject the unit root null in real exchange rates can no longer be ascribed to low power of the tests.

[33] Since we are using the size and power tests with $\alpha = 1.104$ in order to correct for size distortions when $\alpha = 1.00$, we interpret the results in terms of the mix between stationary and unit root countries.

We investigate the hypothesis that the failure to reject unit roots in real exchange rates with panel methods can be explained by the great appreciation and depreciation of the dollar. We extend Perron's (1989) changing growth model to develop univariate and panel unit root tests that allow for three breaks in the slope of the trend function, with the dates of the breaks determined endogenously. The coefficients on the break dummy variables are restricted so as to produce a constant mean prior to the first and following the third breaks. Furthermore, the coefficients are constrained so that the pre and post-break means are equal. These restrictions ensures that rejection of the unit root null in favor of the PPP restricted broken trend alternative is evidence of long-run purchasing power parity.

Even among the class of models which satisfy the PPP restrictions, the range of possibilities for multiple structural changes are enormous. In general, there can be multiple breaks in the intercept, slope, or both, of the trend function. Post-1973 dollar-based real exchange rates, drawn from a nominal flexible exchange rate regime, appear to be better characterized by long swings (slope changes) than by discrete jumps (intercept changes). Although our choice of three breaks is motivated by the rise and fall of the dollar, it also receives statistical support. We use tests developed by Bai (1999) to show that, out of the class of models with up to five slope changes, the strongest evidence of structural change for 13 out of 20 countries is for three breaks.

While the univariate tests do not produce evidence against unit roots in real exchange rates, they provide a classification of the 20 countries into two groups. The real exchange rates of the 15 countries with the smallest root median squared error (RMSE) of the break dates follow the typical pattern associated with the sharp rise and even sharper fall of the dollar in the 1980s, and have breaks that are, on average, less than one year away from the median. The exceptions to the typical pattern, in descending order of their RMSE's, are Japan, Australia, Portugal, Canada, and Greece.

The central result of the paper is that panel unit root tests that account for PPP restricted structural change provide very strong evidence against the unit root hypothesis, and thus evidence of purchasing power parity, for panels of between 11 and 15 typical countries, as well as for the panel of 16 which includes one atypical country. Unit root tests that do not account for structural change provide no evidence of PPP for these panels. For larger panels of 17 to 20 countries, which include more atypical countries, the opposite occurs. Incorporating structural change decreases the evidence of PPP.

Does this constitute evidence of PPP? We conduct simulations to investigate the size and power of the tests, and find that the results are most consistent with the hypothesis that the 15 typical countries are stationary while the others contain a unit root. Furthermore, the results are clearly inconsistent with either more than 16 or fewer than 14 stationary countries. Our conclusions are twofold: First, we find very strong evidence that PPP holds for most of the countries. Second, we find that PPP does not hold for all countries. The delineation among countries is very sharp. PPP holds for those countries, almost all European, that follow the typical pattern of the rise and fall of the dollar in the 1980s. PPP does not hold for the other, almost all non-European, countries that do not follow the typical pattern.

References

Abuaf, N. and P. Jorion, (1990), Purchasing power parity in the long run, Journal of Finance, 45, 157-174.

Andrews, D., (1993), Exactly median unbiased estimation of first order autoregressive/unit root models, Econometrica, 61, 139-165.

Bai, J., (1999), Likelihood ratio tests for multiple structural changes, Journal of Econometrics, 91, 299-323.

Balassa, B., (1964), The purchasing power parity doctrine: a reappraisal, Journal of Political Economy, 72, 584-596.

Bowman, D., (1999), Efficient tests for autoregressive unit roots in panel data, manuscript, Federal Reserve Board.

Breuer, J. (1994), An assessment of the evidence on purchasing power parity, in: J. Williamson, ed., Estimating equilibrium exchange rates, (Institute for International Economics, Washington D.C.)

Breuer, J., McNown, R. and M. Wallace, (2000), Misleading inferences from panel unit root tests with an illustration from purchasing power parity, forthcoming, Review of International Economics.

Campbell, J. and P. Perron, (1991), Pitfalls and opportunities: What macroeconomists should know about unit roots, NBER Macroeconomics Annual, 141-201.

Canova, F., (1997), Testing for heterogeneity in the cross sectional dimension of a panel: A predictive density approach, manuscript, Universitat Pompeu Fabra.

Canzoneri, M., Cumby, R., and B. Diba, (1999), Relative labor productivity and the real exchange rate in the long run: Evidence for a panel of OECD countries, Journal of International Economics, 47, 245-266.

Cheung, Y.-W. and K. Lai, (2000), On cross-country differences in the persistence of real exchange rates, Journal of International Economics., 50, 375-397.

Dornbusch, R., (1976), Expectations and exchange rate dynamics, Journal of Political Economy, *84*, 1161-1176.

Dornbusch, R. and T. Vogelsang, (1991), Real Exchange rates and purchasing power parity, in: J. de Melo and A Sapir, eds., Trade theory and economic reform: North, South, and East, essays in honor of Bela Balassa (Basil Blackwell, Cambridge MA), 3-24.

Elliot, G., Rothenberg, T. and J. Stock (1996), Efficient tests for an autoregressive unit root, Econometrica, 64, 813-836.

Engel, C., (2000), Long-run ppp may not hold after all, Journal of International Economics, 51, 243-273.

Engel, C. and J. Hamilton, (1990), Long swings in the dollar: Are they in the data and do markets know it?, American Economic Review, 80, 689-713.

Engel, C., M. Hendrickson, and J. Rogers, (1997), Intranational, intracontinental, and intraplanetary PPP, Journal of the Japanese and International Economies, 11, 480-501.

Frankel, J. and K. Froot, (1990), Chartists, fundamentalists, and trading in the foreign exchange market, American Economic Review, 80, No. 2, 181-185.

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.

Froot, K. and K. Rogoff, (1995), Perspectives on PPP and long-run real exchange rates in Handbook of International Economics, Volume III, G. Grossman and K. Rogoff, eds., North-Holland, Amsterdam., 1647-1688.

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.

Hegwood, N. and D. Papell, (1998), Quasi purchasing power parity, International Journal of Finance and Economics, 3, 279-289.

Higgins, M. and E. Zakrajsek, (1999), Purchasing power parity: Three stakes through the heart of the unit root null, manuscript, Federal Reserve Bank of New York.

Im, K., Pesaran, H. and Y. Shin, (1997), Testing for unit roots in heterogeneous panels, manuscript, University of Cambridge.

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.

Levin, A. and C.F. Lin, (1992), Unit root tests in panel data: asymptotic and finite-sample properties, UCSD Discussion Paper 92-23, May.

Lothian, J., (1998), Some new stylized facts of floating exchange rates, Journal of International Money and Finance, 17, 29-39.

Lothian, J. and M. Taylor (1997), Real exchange rate behavior, Journal of International Money and Finance, 16, 945-954.

Lumsdaine, R. and D. Papell, (1997), Multiple trend breaks and the unit root hypothesis, Review of Economics and Statistics, 79, 212-218.

Maddala, G.S. and S. Wu, (1999), A comparative study of unit root tests with panel data and a new simple test, Oxford Bulletin of Economics and Statistics, 61, 631-652.

Ng, S. and P. Perron, (1995), Unit root tests in arma models with data dependent methods for the selection of the truncation lag, Journal of the American Statistical Association, 90, 268-281.

Obstfeld, M. and A. Taylor, (1997), Nonlinear aspects of goods-market arbitrage and adjustment: heckscher's commodity points revisited, Journal of the Japanese and International Economies, 441-479.

O'Connell, P., (1998a), The overvaluation of purchasing power parity, Journal of International Economics, 44, 1-19.

O'Connell, P., (1998b), Market frictions and real exchange rates, Journal of International Money and Finance, 17, 71-95.

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

Papell, D., (1997), Searching for stationarity: Purchasing power parity under the current float, Journal of International Economics, 43, 313-332.

Papell, D. and H. Theodoridis, (1998), Increasing evidence of purchasing power parity over the current float, Journal of International Money and Finance, 17, 41-50.

Papell, D. and H. Theodoridis, (2000), The choice of numeraire currency in panel tests of purchasing power parity, forthcoming, Journal of Money, Credit and Banking.

Pedroni, P. (1997), Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis, new results, manuscript, Indiana University.

Perron, P., (1989), The great crash, the oil price shock, and the unit root hypothesis, Econometrica, 57, 1361-1401.

Perron, P., (1997), Further evidence on breaking trend functions in macroeconomic variables, Journal of Econometrics, 80, 355-385.

Rogoff, K., (1996), The purchasing power parity puzzle, Journal of Economic Literature 34, 647-668.

Samuelson, P., (1964), Theoretical Notes on Trade Problems, Review of Economics and Statistics, 46, 145-164.

Stockman, A., (1990), Exchange rates, the current account, and monetary policy, in W. Haraf and T. Willett, eds., Monetary policy for a volatile global economy, (AEI Press, Washington DC), 95-116.

Taylor, M. and D. Peel, (2000), Nonlinear adjustment, long-run equilibrium and exchange rate fundamentals, Journal of International Money and Finance, 19, 33-53.

Wu, S. and J.-L. Wu, (1999), Purchasing power parity under the current float: New evidence from panel data unit root tests, manuscript, SUNY at Buffalo .

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

Vogelsang, T., (1998), Sources of nonmonotonic power when testing for a shift in the trend of a dynamic time series, Journal of Econometrics, 88, Issue 2, 283-299.

Table 1

Structural Change Tests

Country	Number of Breaks	p-values
Australia	4	.022
Austria	3	.006
Belgium	3	.008
Canada	2	.088
Denmark	3	.006
Finland	5	.075
France	3	.015
Germany	3	.015
Greece	3	.014
Ireland	4	.186
Italy	3	.273
Japan	3	.421
Netherlands	3	.008
New Zealand	5	.025
Norway	3	.011
Portugal	3	.062
Spain	3	.098
Sweden	5	.029
Switzerland	3	.046
United Kingdom	4	.131

Table 2
Univariate Restricted Unit Root Tests

Country	α	t-statistic	p-values	k
Australia	-0.131	-2.53	.562	0
Austria	-0.169	-2.47	.587	2
Belgium	-0.179	-2.74	.474	3
Canada	-0.046	-1.31	.922	6
Denmark	-0.255	-3.11	.328	2
Finland	-0.176	-2.80	.449	7
France	-0.304	-4.15	.071	0
Germany	-0.229	-2.73	.477	2
Greece	-0.384	-3.85	.113	4
Ireland	-0.106	-1.64	.863	6
Italy	-0.157	-2.85	.429	0
Japan	-0.147	-2.53	.562	3
Netherlands	-0.242	-3.08	.342	4
New Zealand	-0.200	-2.74	.474	3
Norway	-0.465	-3.01	.365	8
Portugal	-0.238	-3.78	.126	0
Spain	-0.149	-2.41	.614	8
Sweden	-0.267	-3.20	.295	8
Switzerland	-0.242	-2.54	.558	6
United Kingdom	-0.127	-2.15	.717	5

Break Dates and Coefficient

Country	TB1	TB2	TB3	RMSE	γ_1	γ_2	γ_3
Australia	81(II)	85(II)	96(I)	34.64	.018	-.024	.007
Austria	79(IV)	85(I)	87(I)	3.35	.025	-.091	.066
Belgium	80(III)	84(IV)	87(II)	1.12	.036	-.097	.061
Canada	77(III)	86(II)	88(II)	13.46	.004	-.021	.017
Denmark	80(I)	85(I)	87(I)	2.50	.028	-.099	.070
Finland	80(III)	84(IV)	87(I)	1.80	.023	-.065	.043
France	80(III)	84(IV)	87(I)	1.80	.031	-.089	.058
Germany	80(II)	85(I)	87(I)	1.80	.029	-.097	.069
Greece	80(II)	84(IV)	90(III)	12.58	.026	-.046	.020
Ireland	80(III)	85(I)	86(II)	4.50	.019	-.086	.067
Italy	79(IV)	85(I)	86(III)	4.61	.021	-.093	.072
Japan	84(III)	95(III)	96(I)	56.66	-.013	.306	-.293
Netherlands	80(III)	85(I)	87(I)	1.50	.029	-.094	.065
New Zealand	80(IV)	85(I)	87(III)	1.12	.026	-.070	.044
Norway	80(III)	84(IV)	87(III)	1.12	.022	-.067	.044
Portugal	79(III)	84(I)	90(III)	13.72	.032	-.053	.022
Spain	80(III)	84(IV)	87(III)	1.12	.033	-.083	.050
Sweden	80(IV)	84(III)	88(I)	3.35	.035	-.073	.038
Switzerland	79(III)	85(IV)	86(III)	6.10	.021	-.193	.173
United Kingdom	82(I)	84(IV)	87(III)	6.10	.032	-.063	.032

Note: The critical values for the unit root tests are -5.00 (1 percent), -4.30 (5 percent), and -3.91 (10 percent).

Table 3

Panel Unit Root Tests

Countries	Excluded	α	t-statistic	Critical Values			p-values
				1%	5%	10%	
20	None	-0.088	-8.62	-9.92	-9.08	-8.70	.114
19	Japan	-0.083	-7.85	-9.57	-8.84	-8.48	.233
18	Australia	-0.081	-7.51	-9.29	-8.56	-8.21	.248
17	Portugal	-0.083	-7.36	-9.09	-8.34	-7.93	.224
16	Canada	-0.128	-8.96	-8.82	-8.10	-7.74	.008
15	Greece	-0.128	-8.48	-8.36	-7.71	-7.37	.007
14	United Kingdom	-0.131	-8.55	-8.22	-7.54	-7.15	.004
13	Switzerland	-0.123	-8.06	-8.00	-7.28	-6.92	.008
12	Italy	-0.124	-7.61	-7.56	-6.95	-6.61	.009
11	Ireland	-0.126	-7.37	-7.34	-6.66	-6.30	.009

Break Dates and Coefficients

Countries	TB1	TB2	TB3	γ_1	γ_2	γ_3
20	85(III)	91(IV)	94(IV)	-.007	.022	-.015
19	85(IV)	91(IV)	94(IV)	-.007	.022	-.015
18	85(IV)	91(IV)	94(IV)	-.008	.024	-.016
17	85(IV)	91(IV)	94(IV)	-.008	.023	-.016
16	80(III)	85(I)	87(IV)	.026	-.067	.042
15	80(II)	85(I)	87(III)	.024	-.071	.046
14	80(III)	85(I)	87(IV)	.026	-.068	.042
13	80(III)	85(I)	87(IV)	.026	-.068	.042
12	80(III)	85(I)	87(IV)	.026	-.069	.043
11	80(III)	85(I)	87(III)	.026	-.073	.047

No Breaks

Countries	α	t-statistic	Critical Values			p-values
			1%	5%	10%	
20	-0.065	-7.62	-8.79	-8.08	-7.70	.114
19	-0.065	-7.35	-8.57	-7.89	-7.46	.123
18	-0.063	-7.05	-8.41	-7.64	-7.22	.131
17	-0.064	-6.86	-8.25	-7.34	-6.99	.179
16	-0.065	-6.59	-7.75	-7.09	-6.72	.122
15	-0.066	-6.44	-7.44	-6.81	-6.44	.100
14	-0.063	-6.08	-7.34	-6.61	-6.23	.131
13	-0.060	-5.74	-7.08	-6.35	-6.00	.151
12	-0.060	-5.56	-6.74	-6.13	-5.80	.149
11	-0.062	-5.38	-6.58	-5.91	-5.53	.130

Note: The countries that comprise the panel of 20 real exchange rates are listed in Tables 1 and 2. The smaller panels are constructed by sequentially removing the countries listed under “excluded”.

Table 4

Power of Panel Unit Root Tests with Structural Change

Countries	Excluded	Stationary with PPP Restrictions			Unit Root with Structural Change		
		1%	5%	10%	1%	5%	10%
20	None	.704	.849	.905	.229	.408	.502
19	Japan	.734	.862	.920	.262	.427	.519
18	Australia	.720	.869	.917	.245	.406	.491
17	Portugal	.695	.855	.915	.256	.409	.524
16	Canada	.815	.911	.948	.309	.469	.557
15	Greece	.825	.921	.956	.342	.496	.586
14	United Kingdom	.804	.898	.943	.327	.481	.583
13	Switzerland	.796	.905	.941	.286	.462	.573
12	Italy	.792	.899	.939	.275	.413	.500
11	Ireland	.786	.889	.936	.255	.407	.509

Note: The countries that comprise the panel of 20 real exchange rates are listed in Tables 1 and 2. The smaller panels are constructed by sequentially removing the countries listed under "excluded".

Table 5

Power of Panel Unit Root Tests with a Mix of Stationary and Unit Root Countries

Number of Countries	**Number of Unit Root Countries 1% Size**											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
20	.655	.613	.555	.521	.489	.466	.429	.384	.365	.346	.305	.294
19	.656	.608	.586	.562	.519	.465	.436	.425	.378	.362	.361	
18	.634	.571	.530	.501	.432	.415	.389	.336	.321	.307		
17	.620	.564	.539	.470	.410	.386	.347	.298	.295			
16	.727	.691	.598	.544	.513	.418	.380	.372				
15	.776	.681	.626	.596	.505	.478	.462					
14	.688	.611	.552	.442	.397	.369						
13	.677	.619	.479	.422	.387							
12	.670	.474	.402	.362								
11	.514	.428	.386									

Number of Countries	**Number of Unit Root Countries 5% Size**											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
20	.832	.791	.730	.720	.689	.661	.626	.582	.585	.521	.489	.472
19	.817	.783	.763	.753	.711	.661	.628	.597	.543	.534	.501	
18	.785	.745	.714	.680	.631	.582	.579	.527	.481	.474		
17	.812	.762	.727	.662	.600	.600	.507	.499	.486			
16	.864	.823	.757	.699	.676	.594	.550	.531				
15	.877	.793	.752	.727	.640	.604	.578					
14	.810	.740	.705	.593	.555	.525						
13	.831	.788	.643	.588	.541							
12	.805	.625	.532	.499								
11	.672	.584	.539									

Number of Countries	**Number of Unit Root Countries 10% Size**											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
20	.888	.861	.809	.799	.765	.754	.709	.677	.675	.614	.580	.559
19	.878	.842	.831	.811	.797	.749	.722	.698	.635	.620	.589	
18	.851	.823	.793	.762	.712	.682	.664	.609	.572	.562		
17	.883	.843	.823	.769	.731	.721	.625	.615	.590			
16	.908	.883	.819	.774	.748	.685	.625	.609				
15	.915	.851	.812	.792	.701	.675	.646					
14	.864	.819	.778	.684	.636	.616						
13	.887	.851	.726	.677	.624							
12	.872	.609	.572	.562								
11	.745	.668	.614									

Note: The countries that comprise the panel of 20 real exchange rates are listed in Tables 1 and 2. The smaller panels are constructed by sequentially removing the countries in descending order of their RMSE's in Table 2.

Table 6

Power of Panel Unit Root Tests with a Mix of Stationary and Nonstationary Countries

Number of Countries	Number of Nonstationary Countries 1% Size											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
20	.450	.294	.192	.154	.115	.090	.063	.057	.045	.032	.028	.029
19	.483	.290	.254	.177	.136	.092	.080	.062	.051	.041	.031	
18	.385	.296	.182	.137	.077	.058	.050	.048	.039	.028		
17	.544	.311	.208	.118	.085	.064	.042	.040	.029			
16	.408	.258	.151	.101	.073	.047	.034	.028				
15	.478	.234	.161	.120	.090	.066	.059					
14	.340	.193	.144	.078	.049	.040						
13	.374	.210	.100	.054	.052							
12	.340	.138	.076	.052								
11	.218	.106	.080									

Number of Countries	Number of Nonstationary Countries 5% Size											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
20	.619	.451	.328	.289	.223	.175	.138	.119	.102	.078	.080	.064
19	.627	.428	.379	.283	.214	.164	.138	.118	.089	.082	.085	
18	.508	.425	.313	.238	.153	.105	.090	.086	.069	.068		
17	.723	.451	.341	.214	.159	.127	.088	.083	.063			
16	.539	.382	.243	.166	.138	.090	.081	.059				
15	.588	.327	.219	.186	.141	.119	.100					
14	.440	.283	.205	.136	.093	.078						
13	.508	.322	.167	.094	.083							
12	.454	.211	.124	.090								
11	.315	.181	.139									

Number of Countries	Number of Nonstationary Countries 10% Size											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
20	.693	.534	.413	.356	.285	.225	.187	.161	.137	.110	.113	.104
19	.701	.490	.451	.355	.259	.196	.185	.154	.129	.117	.120	
18	.584	.501	.369	.279	.190	.142	.121	.113	.093	.095		
17	.800	.533	.434	.290	.207	.190	.134	.115	.100			
16	.613	.444	.297	.210	.176	.127	.103	.079				
15	.641	.376	.277	.235	.183	.152	.135					
14	.509	.348	.252	.182	.127	.111						
13	.570	.381	.225	.144	.114							
12	.520	.247	.157	.120								
11	.368	.218	.173									

Note: The countries that comprise the panel of 20 real exchange rates are listed in Tables 1 and 2. The smaller panels are constructed by sequentially removing the countries in descending order of their RMSE's in Table 2.

Figure 1: Nominal and Real Exchange Rates

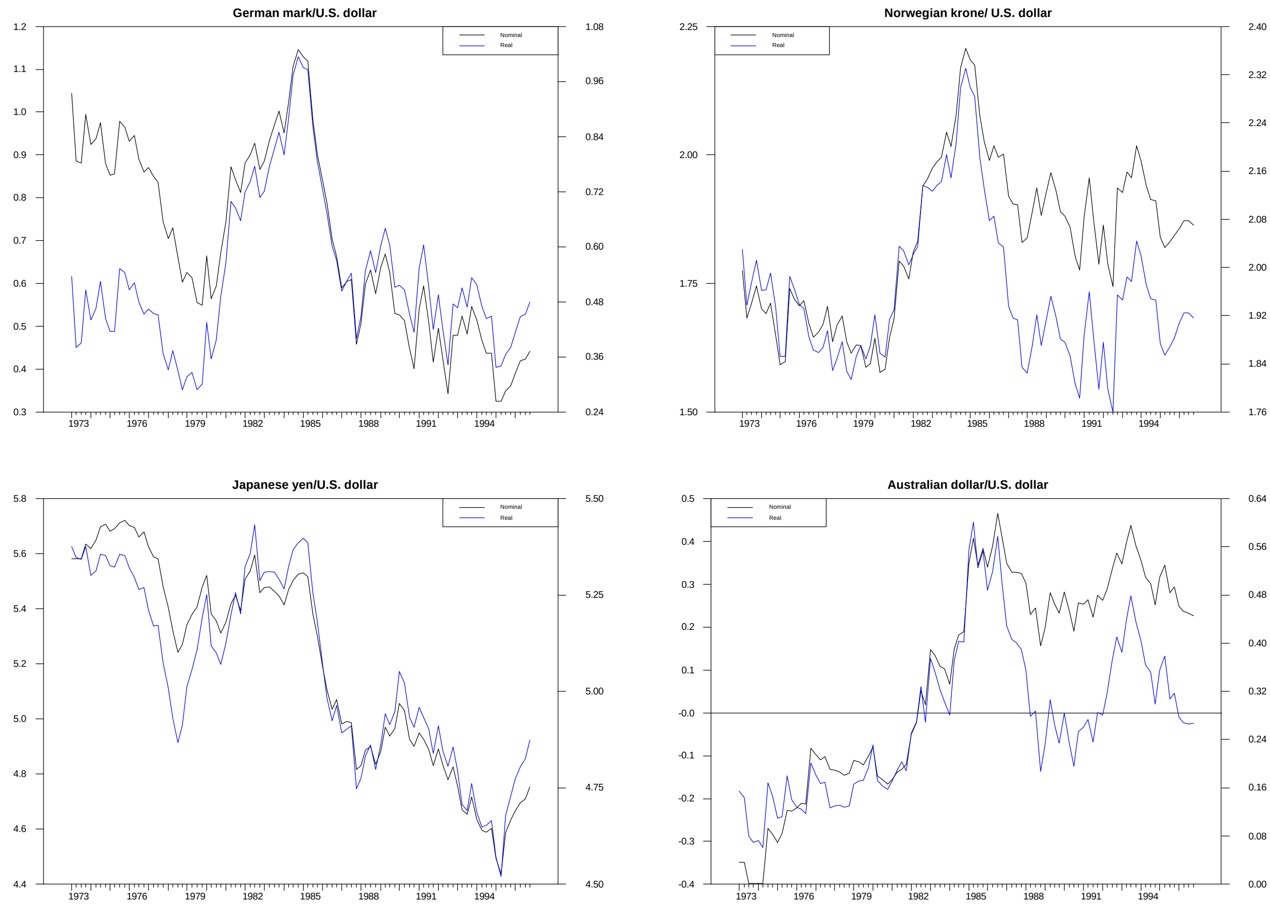

Figure 2: Real Dollar Exchange Rates

"Typical" Countries

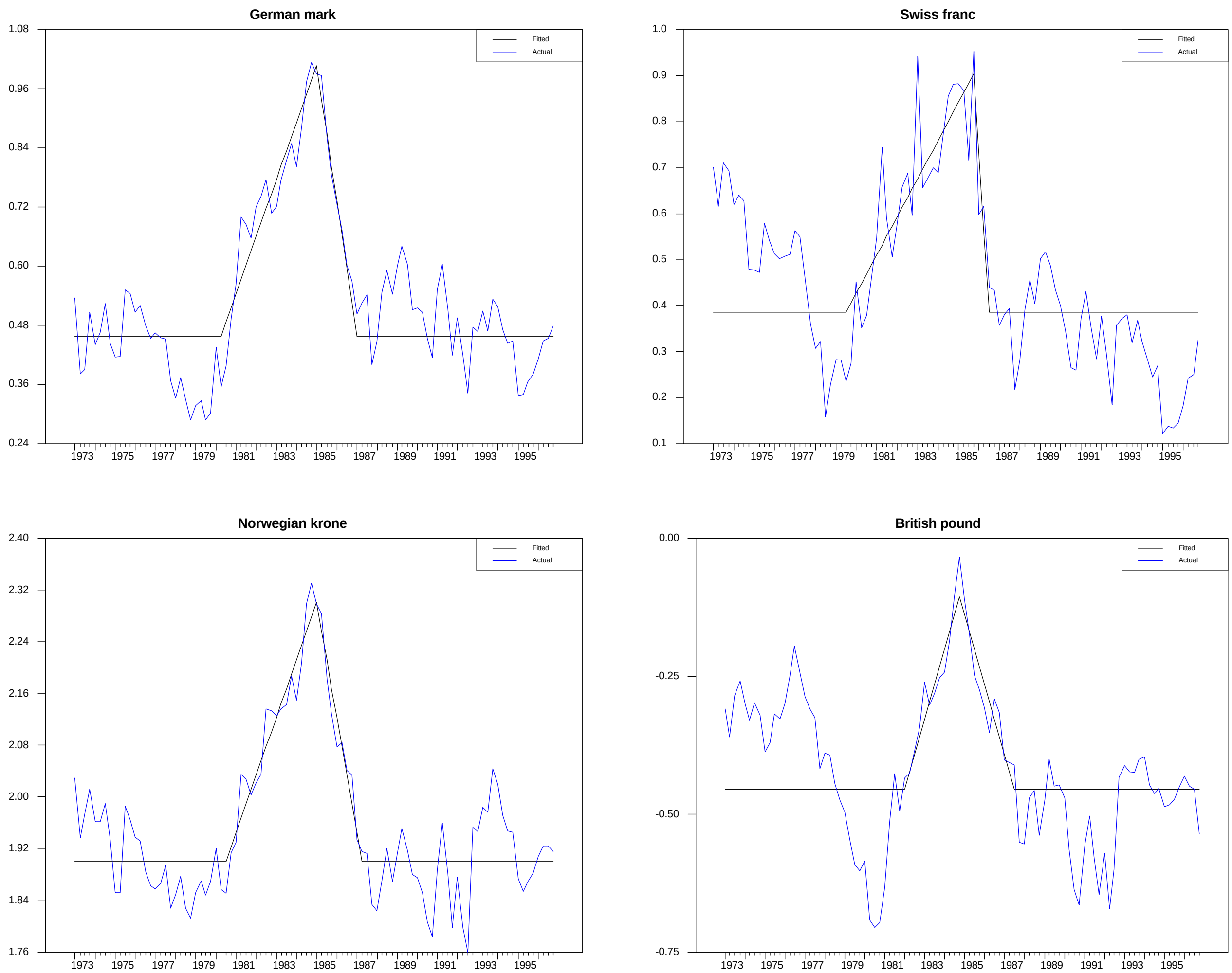

Figure 3: Real Dollar Exchange Rates

"Atypical" Countries