Nonlinear Trends in Real Exchange Rates: A Panel Unit Root Test Approach

by

David O. Cushman
Westminster College, New Wilmington, PA, USA
&
University of Saskatchewan, Saskatoon, SK, Canada

Nils Michael
Health Finance Directorate, The Scottish Government
Edinburgh, Scotland

March 4, 2009

Abstract: We discuss the possibility of deterministic nonlinear trends rather than constants or linear trends as the alternative to unit roots in real exchange rates. To test for this, we extend the panel approach to testing for unit roots by adding nonlinear trends up to time order 5 to Pesaran’s CIPS test (M.H. Pesaran, 2007, “A Simple Panel Unit Root Test in the Presence of Cross-Section Dependence,” *Journal of Applied Econometrics* 22, 265-324), which is then applied to 23 OECD-U.S. real exchange rates, 1974-1998. We employ a bootstrapping procedure that accounts for many data-specific characteristics, data-based lag choice, and multiple applications of the test. The unit root can be rejected at the 0.05 level, and the strongest rejection occurs when quadratic trends are specified. The statistical significance of nonlinear trends is next investigated. The results are somewhat inconclusive. Finally, we compare our specific CIPS test outcomes with simulated CIPS test behavior under various time trend orders. Our CIPS test outcomes are, overall, most consistent with CIPS test behavior under quadratic trends in the real exchange rates.

JEL CODE: F31; F41

KEY WORDS: Real exchange rates; nonlinear-trend stationarity; panel unit root tests; purchasing power parity
1. Introduction

Empirical testing concerning real exchange rates and purchasing power parity (PPP) over the last two decades has generally taken a unit root as the null (PPP does not hold) with stationarity around a fixed mean as the alternative (PPP does hold). Sometimes a linear trend or breaks in mean have been included for the alternative. The results have been mixed. Taylor and Taylor (2004) discuss the variety of strategies that have been employed so far to lessen the ambiguity. These include use of longer estimation periods, panel tests across many exchange rates, and tests with nonlinear adjustment to the mean.

Taylor and Taylor (2004) go on to suggest that another strategy be employed, which is to account for the possibility that real exchange rates may follow nonlinear trends. That is, the means to which real exchange rates are tending to revert may themselves gradually evolve in a nonlinear fashion. Taylor and Taylor (2004) suggest that this could reflect fluctuating net international investment positions or changing productivity differentials. Although nonlinear trends contradict the usual definition of PPP, the relationship may approximate PPP and the existence of a meaningful equilibrium real exchange rate is of interest in any event. If nonlinear trends do characterize real exchange rates, their omission thus far in the various tests might have contributed to the ambiguous results.

Cushman (2008) applies the nonlinear trend idea to a number of individual OECD real exchange rates for the recent floating period, and in some cases it appears that the unit root null can be rejected with nonlinear trend stationarity being the most likely alternative. The statistical significance, however, is not generally strong. In this paper we therefore investigate the possibility of nonlinear trends in 23 U.S. dollar-OECD real exchange rates using a panel approach, which should have more power. The specific panel test employed is the CIPS test of Pesaran (2007). The estimation period is the now classic floating period from the beginning of generalized floating until the introduction of the euro, 1974-1998. As is typical, our real exchange rates are constructed using consumer price indexes.

To determine the significance levels of the test results, we employ a bootstrapping procedure that accounts for many data-specific characteristics, data-based lag choice, and multiple applications of the test. We find that the CIPS test does not reject unit roots at the 0.05 level with only means specified. With linear trends specified, a borderline rejection occurs. And with quadratic trends specified, a clear rejection occurs. The test specified with more complicated nonlinear trends does not reject. Then, our p-value computed to adjust for multiple tests also supports the conclusion that unit roots should be rejected, and we believe this conclusion is relatively reliable because of our conservative approach to p-values.

The clear rejection with quadratic trends and not with means or linear trends suggests that nonlinear trends are important. To augment this evidence, we then run tests on the statistical significance of the nonlinear trends themselves, again following our conservative bootstrapping strategy. The results are somewhat supportive of the presence of nonlinear trends but are not unambiguously significant at the 0.05 level. As a final approach to evaluating the likelihood of nonlinear trends, we return to the CIPS test and investigate its behavior when such trends are
present. Our CIPS results turn out to be, overall, more consistent with quadratic trends in the real exchange rates than with constants, linear trends, or more complicated nonlinear trends.

2. Real exchange rates and nonlinear trends

We now provide some motivation for the possible existence of nonlinear trends in real exchange rates. The basic idea is that there are well-known factors that can cause deviations from PPP and hence fluctuations in real exchange rates, and that some of these factors might follow nonlinear trends. Camarero (2008) provides a summary that leads to a composite model with six real factors affecting real exchange rates. With the addition of two other factors not in Camarero (2008), the model is:

\[
q_t = f[(R_t - R^*)_t, (a_t^T - a^*_t^T), (a_{NT}^T - a^*_{NT}N^T), (a_t^D - a^*_t^D), (g_t - g^*_t), dta_t,
(d_t - d^*_t), (y_t - y^*_t)].
\]

(1)

The first term in Camarero’s (2008) model is the real interest differential, from Meese and Rogoff (1988). Its presence is based on the relationship of real exchange rate changes to the real interest differential given random-walk real shocks and uncovered nominal interest parity. The second term allows changes in relative traded-good sector productivity between the two countries, and the third term allows this for non-traded goods. The two terms capture the well known Belassa (1964) and Samuelson (1964) effect. The fourth term allows relative productivity changes between countries in the distribution sectors for traded goods, from Devereux (1999) and MacDonald and Ricci (2005). Although the law of one price may hold the wholesale tradable goods prices equal between countries, a change in the productivity of one country’s distribution sector can nevertheless affect the retail price, thus impacting the real exchange rate. This can help account for Engel’s (1999) conclusion that most fluctuations in real exchange rates seem to involve changes in the relative prices of traded goods rather than the Balassa-Samuelson effect. The fifth term is the differential of real government expenditure net of taxes and derives from the traditional Mundell (1961, 1962) and Fleming (1962) ISLM model (as in McCallum, 1996, chapter 6).\(^1\) Real fiscal expansions in a given country appreciate that country’s currency in real terms.\(^2\) The sixth term is the difference in the two countries’ net foreign asset positions (relative to GDP), and can be derived from Hooper and Morton (1982) and various portfolio balance exchange rate models (see Hallwood and MacDonald, 2000). Growing foreign asset positions appreciate a country’s currency.

We add two more terms not in Camarero’s (2008) model. The first follows from recognizing that portfolio balance models generally include government debt in addition to net foreign assets. Growth of government debt depreciates the country’s currency. This adds the seventh term in equation (1). The second term we add, and the last in equation (1), is the real output differential. It derives from the ISLM model. Exogenous real output growth depreciates the given country’s real exchange rate (McCallum, 1996, chapter 6).

---

\(^1\) This abstracts from possible balanced budget effects.

\(^2\) Fiscal policy is also present in more modern macro models. See, for example, Obstfeld and Rogoff (1995) and Ganelli (2005). However, in these papers purchasing power parity is assumed and so the findings are not directly relevant to the study of real exchange rates.
Although the steady-state real exchange rate is constant in the absence of various real fluctuations, over time there are likely to be either sudden or gradual non-offsetting changes in the real factors in equation (1). Furthermore, adjustment toward the new steady state is unlikely to be instantaneous. Therefore, the time series properties of the real factors and the nature of adjustment will determine the behavior of the real exchange rate. Some of the net fluctuations could be nonstationary in either a stochastic or deterministic manner, and so there will be stochastic or deterministic trends in real exchange rates. A detailed examination exceeds the scope of the paper, but several possibilities suggestive of nonlinear trends can be highlighted.

One possibility would stem from the Balassa-Samuelson effect. Consider first that a constant productivity growth rate differential between traded and nontraded goods gives rise to a linear trend in the real exchange rate (see Devereux, 1999). The upward trend in the real Japanese yen from the 1970s to the 1990s has been cited as an example (Marston, 1991, Devereux, 1999). Now suppose that the productivity growth rate differential changed gradually. Then a nonlinear trend in the real exchange rate becomes plausible. Devereux (1999) shows how the presence of a distribution sector for tradables can reverse the direction of the real exchange rate’s trend that is predicted by the Balassa-Samuelson effect. Moreover, he explicitly derives a case where a one-time rise in the relative productivity growth rate in traded goods generates a nonlinear trend during the adjustment to the new steady state linear trend in the real exchange rate.\(^3\)

For another possibility, consider the final variable in equation (1), \((y_t - y_t^*)\). Real outputs have often been modeled as possessing stochastic trends. But some argue otherwise. For example, Papell and Prodan (2004) conclude that, with several temporary and endogenously determined structural breaks accounted for, the log of U.S. real GDP is linear-trend stationary. The breaks do not occur during the post-Bretton Woods period. If this conclusion can be extended to other countries, then in equation (1) \(y_t\) and \(y_t^*\) are apparently linear-trend stationary during our estimation period. However, the apparent linear trends might actually be approximations to nonlinear trends in log real GDPs, a possibility mentioned by Stock and Watson (1999).

Such nonlinear trends can be motivated by the long-run slowdown in productivity growth as an economy approaches the steady state from an initial position of capital scarcity. From results in Barro and Sala-i-Martin (1992) and the definition of effective labor therein, it can be computed that real GDP would possess the following nonlinear trend:

\[
y_t = l_0 + (x + n)t + \hat{y}_0 e^{-\beta t} + \hat{y}^e (1 - e^{-\beta t}),
\]

where \(l_0\) is the log of labor in initial time period 0, \(x\) is the rate of labor-augmenting technological change, \(n\) is the rate of labor growth, \(\hat{y}_0\) is the log of output per unit of effective

---

\(^3\) In his example, the change in the traded goods productivity growth rate is abrupt, leading to a sudden jump in the real exchange rate before it embarks on its nonlinear path toward the new steady state. However, it is reasonable to suppose that such productivity enhancements might be introduced gradually, leading to an entirely smooth nonlinear real exchange rate path.
labor in time period 0, \( \hat{y}^x \) is the log of output per unit of effective labor in the steady state, and \( \beta \) is an adjustment speed parameter. Suppose equation (2) also holds for the foreign country, with the exception that \( \hat{y}_0 \neq \hat{y}_0^* \). Then

\[
(y_t - y_t^*) = (\hat{y}_0 - \hat{y}_0^*)e^{-\beta t}.
\]  

(3)

When inserted in equation (1), this provides another source of nonlinear trends in the real exchange rate.\(^4\)

Finally, consider the role of real fiscal balances and their interaction with government debt. In general, governments might have targets for the ratios of fiscal balance and debt to real GDP that remained relatively constant. However, a number of countries in our application were involved in the Maastricht convergence criteria that required significant adjustment to the ratios. Accounting for a variety of adjustment costs, Dworak, Wirl, Prskawetz, and Feichtinger (2002) examine optimal time paths for the ratios. The paths are nonlinear. Consequently, if real GDPs are trend stationary (linear or nonlinear), then the government fiscal balances and debt levels would be nonlinear trend stationary, possibly generating nonlinear trends in real exchange rates.

3. The existing panel-unit-root-test literature and how the present paper fits in

To show how our paper fits in with past empirical work, we now present a summary of relevant literature. To keep the length manageable, our discussion is restricted primarily to papers applying panel unit root tests to post-Bretton Woods real exchange rates.

In large part, the literature has been a progression of attempts to exploit the power of the panel approach while avoiding numerous econometric problems that can distort the results. How many exchange rates are used, which ones are used, and which country is designated the numeraire or base? In the given panel test, are the individual real exchange rates all constrained to adjust to the long-run mean at the same speed under the alternative hypothesis of stationarity? How is serial correlation dealt with, and, in particular, are serial correlation processes allowed to differ among the individual exchange rates? Does the test account for contemporaneous (cross-section) correlation? Are any deterministic terms besides constants are included in the tests, that is, are there trends or breaks? Is the adjustment to the long-run mean a linear or nonlinear process under the alternative hypothesis?

Once the tests have been computed, what method is used to get critical values or p-values? Is the distribution used the asymptotical one that applies to the test (or, otherwise, merely adjusted for the given sample size)? Or is a data-based method used (i.e., bootstrapping)? If the latter, are the serial correlation processes accounted for, are the errors allowed to be non-normal, and are they allowed to be contemporaneously correlated? Is the effect of data-based lag selection accounted for?

---

\(^4\) The trend will be more pronounced the greater the initial disparity in output per unit of effective labor, and so would most likely apply to the real exchange rates of the relatively low-income countries in our panel.
To help summarize the development of the literature, we have listed a large number of the relevant papers in Table A in the Appendix. The table summarizes how each paper has addressed the various questions and our reading of its general conclusion regarding unit roots. Usually, panel-test papers have focused on the alternative hypothesis of constant means rather than including trends or breaks. The U.S. has usually been the base country for the exchange rates.

The work can be divided into several generations. The first used the Levin and Lin (1993) (LL) approach and clearly rejected unit roots in favor of mean stationarity. But the applications mostly employed restrictive serial correlation specifications, used non-data-based sampling distributions (which were prone to generate size distortion), and overlooked contemporaneous correlation, which is significant for exchange rates (and is another source of serious size distortion). These problems were forcefully critiqued by Papell (1997) regarding serial correlation and the need for data-based sampling distributions, and by O’Connell (1998) regarding contemporaneous correlation. Papell (1997) also argued that data-based lag selection was another source of size distortion and should be accounted for.

There consequently followed a second generation of papers that accounted for contemporaneous correlations by using GLS estimation, and that adjusted for size distortion by generating data-based sampling distributions. This generation of papers rejected unit roots less often than the first. The sensitivity of the results to the choice of base countries was also pointed out (the most thorough example being Papell and Theodoridis, 2001).

Yet another generation of papers relaxed the equal-adjustment-speed restriction. In so doing, Taylor and Sarno (1998), Wu and Wu (2001), Choi (2001), and Smith, Leybourne, Kim, and Newbold. (2004) all rejected unit roots. Taylor and Sarno (1998) continued to use GLS to control for contemporaneous correlation. The remaining papers, however, used the Im, Pesaran, and Shin (1997, 2003) (IPS) test, the Maddala and Wu (1999) (MW) test, or variations, which were not specified to control for contemporaneous correlation. Except for Choi (2001), these papers relied instead on bootstrapping to try to avoid the size distortion otherwise likely to occur. Pesaran (2007) then modified the IPS test to account for contemporaneous correlation (the CIPS test) and was unable to reject unit roots in his real exchange rate panel.

Meanwhile, efforts in the vein of the second generation continued (Papell, 2006, Lopez and Papell, 2007, Lopez, 2008). However, these papers were less concerned with simply rejecting the unit root than with the how the decision related to the length of the estimation period, the cycle of real exchange rate appreciations and depreciations, and the composition of the panel.

---

5 All the papers we list in the table take unit roots as the null. We are aware of one panel paper that takes mean stationarity as the null, Kuo and Mikkola (2001), who are unable to reject this null, supporting PPP.
7 Abuaf and Jorion (1990) were ahead of their time in their efforts to account for several of these issues.
8 In addition to Papell (1997) and O’Connell (1998), the list includes Jorion and Sweeney (1996), Papell and Theodoridis (1998), Papell and Theodoridis (2001).
9 They also employed the Johansen cointegration approach to test the null hypothesis of 3 or fewer of their 4 real exchange rates stationary against the alternative of all 4 stationary. This hypothesis was rejected.
A final generation of papers addressed the possibility of nonlinear adjustment toward the mean under the alternative hypothesis. The idea is that the fraction of the gap closed in a given time period could be greater for large gaps than for small gaps, rather than constant as in standard unit root tests. Taylor, Peel, Sarno (2001) modified the approach of Taylor and Sarno (1998) to account for this idea, while Cerrato, de Peretti, and Sarantis (2008) modified the CIPS test. Both papers rejected unit roots, although Cerrato et al. (2008) did not employ data-based approaches to obtain their critical values.

Thus we find that the more recent generations of papers generally conclude that unit roots should be rejected in favor of mean stationarity. This state of affairs may nevertheless not be definitive. First, although the more conservative, data-based methods of computing p-values have become standard, no single paper accounts for all the issues we have included in Table A. For example, non-normality is usually not present in the data-generating processes. Yet, it is generally believed that exchange rate changes are non-normal. And in the two papers that do allow for non-normality, the effect of data-based lag selection is absent. Thus, the conclusion of mean stationarity might not survive testing that accounts for these problems.

The second reason that the mean-stationarity conclusion might not be definitive is that, even if the unit root null should be rejected, the correct alternative might not be mean stationarity. Panel tests that are specified for means might also have power against other alternatives, such as the nonlinear trends that we investigate.

Therefore, our basic idea is to add nonlinear deterministic trends to a panel unit root test. We also wish to address as many of the econometric issues as possible so that any conclusions are as reliable as possible. We address the issue by first choosing a test that specifies both heterogeneous serial correlation processes and contemporaneous correlation. Pesaran’s (2007) CIPS test. Then we go on to apply a bootstrap procedure that allows for specific lag parameter values, non-normality, heterogeneous serial correlation, contemporaneous correlation, and data-based lag selection. Furthermore, the form of the possible nonlinear trends is uncertain and so we apply the CIPS test with several different time-trend specifications. To avoid being misled by a chance rejection among multiple applications of the test, we go on to compute a joint p-value for the multiple test outcomes.

It is unfortunately not possible within the confines of a single paper to consider all specification issues. For example, in choosing the CIPS test we do not constrain the adjustment speeds to the long-run equilibrium to be equal. Thus we implicitly accept the advice of Im, Pesaran, and Shin (1997, 2003) and Maddala and Wu (1999). However, Papell and Theodoridis (2001) and Lopez (2008) argue in favor of the constraint. In particular, if the constraint is true

---

10 A classic paper on this is Westerfield (1977). In our own case, the application of the Jarque-Bera test, as modified by Doornik and Hansen (1994), rejects normality of the residuals from augmented Dickey-Fuller equations at the 0.05 level for 11 of the 23 real exchange rates.

11 A potential econometric problem that has no often been considered in the literature under review is heteroskedasticity. In the present case, an ARCH(4) test (Engle, 1982) applied to the residuals from augmented Dickey-Fuller equations generates no rejections of homoskedasticity at the 0.05 level. The White (1980) test with squares and cross products generates only 2 rejections. And a simple F test that compares the residual variance in the first half of the sample with that in the second half generates no rejections. We conclude that it is reasonable to assume homoskedasticity for our real exchange rate set.
under the alternative, applying it naturally leads to more power. They also argue that a rejection with the constraint applied has the clearer interpretation that all the real exchange rates are stationary. In contrast, rejections from the IPS, MW, and CIPS tests definitely include the possibility that only some exchange rates are stationary. Yet, a rejection with the constraint applied does not actually rule out the latter possibility, because such a test will still have power even if only some real exchange rates are stationary. In any event, we leave the addition nonlinear trends to an adjustment-speed-constrained test for future work.

We also do not consider nonlinear adjustment, again leaving for the future the joint consideration of nonlinear adjustment and nonlinear trends. Sollis (2008), however, concludes that structural change might actually be more important than nonlinear adjustment. Nonlinear trends are a way to model structural change.

4. Pesaran’s test and the addition of nonlinear trends

Pesaran’s CIPS test is an extension of the IPS test of Im, Pesaran, and Shin (1997, 2003). The null hypothesis is a unit root (with or without drift, depending on deterministic terms included) for all the individual time series. The alternative is a stationary process for at least one of the individual series.

The CIPS test statistic is the mean of the CADF statistics for the individuals. The “C” in the names stands for “cross-sectionally augmented.” The test follows from the augmentation of the standard ADF test with lagged levels and first-differences of the cross-section averages of the individual series. In the case of an intercept only, the CADF statistic is the $t$-ratio for $b_i$ in the regression:

$$\Delta y_{it} = a_i + b_i y_{i,t-1} + c_i \bar{y}_{t-1} + \sum_{j=0}^{p} d_j \Delta \bar{y}_{t-j} + \sum_{j=1}^{p} \delta_j \Delta y_{i,t-j} + e_{it},$$

which is eq. (54) in Pesaran (2007). Pesaran goes on to extend the equation to include a linear trend, and here we additionally add nonlinear polynomial time terms. Our earlier theoretical discussion of nonlinear trends in real exchange rates suggests that there could be many possible functional forms for the trends. Our assumption is that polynomial time functions can approximate them. Therefore, the deterministic part of the regression, $a_i$, is redefined as:

$$a_i = a_{0,i} + a_{1,i} t + a_{2,i} t^2 + \ldots + a_{q,i} t^q .$$

In the discussion that follows, the “time order” value refers to the highest exponent, $q$, on $t$ that is included in the test.

---

12 We are using exactly the same generic notation here as in Pesaran (2007), and so the various letters here, other than $t$, have no correspondence to those used in equations (1)-(3).
5. Empirical Results: Test statistics and p-values under the unit root null

We apply the CIPS test to three exchange rate panels. The primary panel consists of 23 OECD-U.S. exchange rates.\(^\text{13}\) The next panel consists of the subset of 17 OECD-U.S. exchange rates examined in Pesaran (2007) and Smith, Leybourne, Kim, and Newbold (2004). The purpose of this sub-panel is to enable us to compare our results with those in Pesaran (2007), who uses the same test (obviously) but without nonlinear trends. The final panel is simply the remaining 6 OECD-U.S. exchange rates.\(^\text{14}\) The data are quarterly and the time period is the much studied post-Bretton Woods floating period from its approximate beginning until the introduction of the euro: 1974:1 through 1998:4.\(^\text{15}\) Pesaran (2007) finds considerable contemporaneous correlation among his 17-exchange-rate panel. His CIPS test is thus quite appropriate for the present paper.

The point of the paper is to allow for nonlinear deterministic trends in the alternative hypothesis, but the actual time order to specify in the test is unclear. Too low a time order will tend to give low power if the actual trends are best approximated by a higher time order. And too high a time order in the test will also tend to give low power if the actual trends are best approximated by a lower time order. A very high time order would also violate Occam’s razor, yet a curve that looked simple to the eye might require a high time order to approximate it. Faced with this, we run the CIPS test for all time orders up to 5.

The lag order for each CADF exchange rate equation is determined by applying a version of the recursive test-down of Ng and Perron (1995). Starting with a maximum lag order of 8, the joint significance of \(\Delta y_{t-8}\) and \(\Delta y_{t-8}\) is tested at the 0.10 level. If not significant, \(\Delta y_{t-8}\) and \(\Delta y_{t-8}\) are dropped, the equation re-estimated, and the joint significance of \(\Delta y_{t-7}\) and \(\Delta y_{t-7}\) is determined. The test-down stops when the joint test is significant, giving the lag order choice. Note that the lag orders of the equations are unlikely to all be the same. The resulting CIPS statistics for all panels and test time orders are in Table 1.

\(^{13}\) This list is approximately the same as those in about half the papers in Table A. It specifically consists of the U.S. dollar exchange rates with the 23 OECD countries with distinct currencies (thus, not Luxembourg) that have been members over the entire floating period, with the addition of Mexico, which joined the OECD in 1994.

\(^{14}\) The 17 exchange rates are for Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, the Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, and the U.K. The remaining 6 are for Greece, Iceland, Ireland, Mexico, Portugal, and Turkey. The data are from the IMF’s *International Financial Statistics*. Some gaps in the CPI data for Iceland are filled in with values from the Central Bank of Iceland’s *Economic Statistics*.

\(^{15}\) After 1998, the nominal exchange rates of the euro countries with the U.S. become identical, so that the only source of variation in the real exchange rate is relative price variation.
Table 1: The CIPS statistics

<table>
<thead>
<tr>
<th>Exchange rates</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
</table>

For an initial impression of statistical significance, we simulate sampling distributions for each number of countries and time order with sample size 100 using driftless random walks with i.i.d. $N(0,1)$ errors and applying the CIPS test with no lagged differences. The p-values from this exercise are in Table 2. Most of the tests appear to be significant at the 0.05 level. Meanwhile, the fail-to-reject conclusion for the panel of 17 at time order 0 is in agreement with the conclusion in Pesaran (2007) for the same panel.

Table 2: CIPS p-values (random walk DGPs and no lagged differences in test)

<table>
<thead>
<tr>
<th>Exchange rates</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0.025</td>
<td>0.022</td>
<td>0.014</td>
<td>0.030</td>
<td>0.29</td>
<td>0.21</td>
</tr>
<tr>
<td>17</td>
<td>0.38</td>
<td>0.091</td>
<td>0.005</td>
<td>0.16</td>
<td>0.021</td>
<td>0.37</td>
</tr>
<tr>
<td>23</td>
<td>0.026</td>
<td>0.003</td>
<td>0.000</td>
<td>0.036</td>
<td>0.001</td>
<td>0.007</td>
</tr>
</tbody>
</table>

Note: P-value significance is highlighted as follows. Values of 0.100 or less are given to 3 decimal places. Values of 0.050 or less are additionally boldface-italic. Values of 0.010 or less are additionally underlined. There are 20,000 replications for each distribution.

Size distortion, however, could be present. For one thing, the relatively good size results reported by Pesaran (2007) for time orders 0 and 1 might not apply to higher time-ordered tests where more time parameters are estimated. Thus, as promised above, we turn to bootstrapping to address the concerns.

First, the lag order of our data-based data generating process (DGP) for each exchange rate must be decided. It seems reasonable to use the AR lag choices made by the test-down procedure as applied to the actual data. But these lag choices are not always the same across the 6 test time orders for a given exchange rate. Yet, we want to examine joint test performance across all the time orders specified, so the same DGP and resulting simulated unit root data sets

---

16 The number of replications for each distribution is 20,000 and the resulting critical values for time orders 0 and 1 are almost identical to those in Pesaran’s (2007) Table II (b) and (c).

17 The specific CIPS value differs, however, for two reasons. First, unlike us, Pesaran (2007) applies the same lag order to all CADF equations (he tries lag values of 1 through 4). Second, his starting date of 1974:1 applies to the estimation period and does not include the initial observations for lags. In the present paper, however, it is the data set that starts in 1974:1, so the estimation periods start later because of lags.
are needed for all tests. Therefore, for each exchange rate we use the median lag choice from the test-down choices across the 6 time orders.

The bootstrapping then proceeds as follows. The first difference equation with the selected lag order is estimated for each of the 23 exchange rates. The estimated residual series for the 23 equations are combined into a matrix (after dropping the early observations for which residuals do not exist in the longer-lag models, and after rescaling and centering). The estimated equations and residuals are then used to recursively create new data series, where the errors for the 23 exchange rate equations for each time period are those from a randomly chosen row of the estimated residual matrix. This is a way to incorporate error nonnormality and contemporaneous correlation present in the original data. The new series are initialized using the first few observations from the actual data.\(^{18}\)

The CIPS tests for all 6 time orders is then applied to each simulated set of 23 exchange rate series. In each set, the lag order for each CADF exchange rate equation in each CIPS test is chosen with the test-down. This is to account for size distortion from having to estimate the lag order as was done with the actual data. The resulting simulated sampling distributions lead to p-values that Papell (1997) calls “exact” p-values. These are in Table 3 and are from the unit root DGPs both with and without (constant) drift, with 20,000 replications of each case.\(^{19}\)

---

\(^{18}\) Using lag orders \(p_i\) from the test-down results, we first estimate \(y_{it} = y_{i, t-1} + \delta_{i} + \sum_{j=1}^{8} \hat{\delta}_{i,j} y_{i, t-j} + \hat{e}_{it}\) for the 23 exchange rates, \(i = 1\) to 23, and the full data set (before loss to lags), \(t = 2\) to 100 with first differences. Given the possibility of 8 lags in first differences, it is the estimated residuals for \(t = 10\) to 100 that are sure to exist for every exchange rate. We form the resulting 91 x 23 residual matrix \(E\). We then build each simulated data set recursively by using the 23 estimated exchange rates equations, replacing the \(\hat{e}_{it}\) with a randomly selected row from \(E\). The first simulated data period is 11, with the actual data from periods 1 to 10 used as necessary for the initial lagged values. Subsequently, newly simulated values become lagged values.

\(^{19}\) The CIPS tests with time orders 0 and 1 allow for no drift and for constant drift (respectively) in their nulls, which are what we specify in our DGPs. However, higher time-ordered CIPS tests allow for linear or nonlinear drift in their nulls. We have not, however, simulated such additional null distributions, because it would involve considerable additional time and probably make little difference. Cushman (2008) found that including nonlinear drift in the null DGPs for bootstrapping DF-GLS tests with nonlinear time terms (applied to individual real exchange rates) made no noticeable difference to the resulting p-values.
Table 3: Bootstrapped CIPS exact p-values

<table>
<thead>
<tr>
<th>Exchange rates</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0.062</td>
<td>0.090</td>
<td>0.094</td>
<td>0.16</td>
<td>0.53</td>
<td>0.48</td>
</tr>
<tr>
<td>6</td>
<td>0.062</td>
<td>0.089</td>
<td>0.095</td>
<td>0.17</td>
<td>0.53</td>
<td>0.48</td>
</tr>
<tr>
<td>17</td>
<td>0.33</td>
<td>0.19</td>
<td>0.065</td>
<td>0.33</td>
<td>0.17</td>
<td>0.56</td>
</tr>
<tr>
<td>17</td>
<td>0.38</td>
<td>0.19</td>
<td>0.057</td>
<td>0.32</td>
<td>0.17</td>
<td>0.55</td>
</tr>
<tr>
<td>23</td>
<td>0.060</td>
<td><strong>0.050</strong></td>
<td><strong>0.031</strong></td>
<td>0.23</td>
<td>0.080</td>
<td>0.18</td>
</tr>
<tr>
<td>23</td>
<td>0.079</td>
<td><strong>0.050</strong></td>
<td><strong>0.028</strong></td>
<td>0.22</td>
<td>0.081</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Note: For each number of countries, the first row results from including drift in the DGP while the second row results from not including drift. P-value significance is highlighted as before. The drift and no-drift values for each set of 6 tests each come from 20,000 bootstrapped replications of the 23 exchange rates.

The p-values in Table 3 give a considerably more conservative impression about rejecting the unit root than those in Table 2. At the 0.05 level, only the 23-country panel at time order 2 gives clear rejections, with the linear trend model on the borderline. The conclusion suggested at this point is that at least some of the exchange rates are deterministic trend stationary and, moreover, least some of the trends gradually evolve in an approximately quadratic manner. The failure of the smaller panels and higher ordered tests to reject could be attributable to lower power in smaller panels and more parameters to estimate with higher time orders.

In our procedure, 6 CIPS tests are applied to a given panel. But, when more than one test is applied, the chances of finding a rejection in one or more tests when applying level $\alpha$ to the individual tests will likely be different from $\alpha$. For example, if 6 independent tests were applied, the probability of finding one rejection at the 0.05 level under the null would considerably higher than 0.05: $P(r = 1) = 0.232$. But the probabilities of larger numbers of rejections are much smaller than 0.05: $P(r = 2) = 0.031$ and $P(r = 3) = 0.002$, with larger numbers of rejections having essentially zero probability. On the other hand, if the 6 tests were almost perfectly positively correlated, the probabilities are rather different: the probabilities of one to five rejections would be close to zero and the probability of 6 would be close to 0.05. This suggests that, to fully assess the evidence against the unit root in Table 3, some adjustment should be made and that it would benefit from incorporating the correlations among the tests. Our bootstrapped distributions implicitly reflect such correlations, because the CIPS tests for all time

---

20 If we were to use the 0.10 level, the 6-exchange-rate sub-panel would reject with three CIPS test time orders and 23-exchange-rate sub-panel would reject with four. The 17-exchange rate panel would reject using time order 2 only. We might then conclude that the failure of the 17-exchange-rate sub-panel to reject for the other time orders could be attributed to its not containing any stationary real exchange rates. But it seems unlikely that we would get a stronger rejection by enlarging the panel from 6 to 23 if none of the added exchange rates were stationary.
orders were applied to the same sequence of simulated data sets. Therefore, we can use the distributions to estimate a joint p-value for the six tests for each panel of exchange rates.

Our joint approach is based on the probability under the unit root null of observing some number of rejections, \( x \), or more at some significance level, \( y \), among the six tests. However, there are numerous \( x \) and \( y \) values to choose from. We resolve this by considering all possible values of \( x \) (in the present case \( x = 1 \) to \( 6 \)) and their corresponding significance levels. Let \( P_x \) be the (ordered) p-values. For each \( x \) the bootstrapped distributions can be used to compute \( P(r \geq x \mid P_x) = \beta^x \). Then we take the most significant (the minimum) of the observed \( \beta^x \) values, \( \beta^*_{\min} \), and compute the value of \( P_x \) (call it \( P_{x,\min} \)) required to generate \( \beta^*_{\min} \) for each of the other five \( x \) values. The joint p-value is the probability of getting \( \beta^*_{\min} \) for any of the six \( x \) values:

\[
P((r \geq 1 \mid P_{1,\min}) \cup (r \geq 2 \mid P_{2,\min}) \cup \ldots \cup (r = 6 \mid P_{6,\min}))
\]

For example, for the panel of 23 exchange rates (with drift in the DGP), the lowest individual p-value in Table 3 is \( P_1 = 0.031 \) and from the bootstrapped distributions the probability of 1 or more rejections at the 0.031 level is \( \beta^1 = 0.136 \). However, it turns out that a finding of 6 rejections at the 0.234 level, which also occurs in Table 3, has the lowest probability, 0.0075, and thus \( \beta^6 = \beta^*_{\min} = 0.0075 \). But just how significant, in fact, is this? There are five other potential rejection outcomes that individually also have a probability of 0.0075: 1 or more rejections at a \( P_{1,\min} \) value of 0.0014, 2 or more rejections at a \( P_{2,\min} \) value of 0.008, 3 or more rejections at a \( P_{3,\min} \) value of 0.025, 4 or more rejections at a \( P_{4,\min} \) value of 0.052, and 5 or more rejections at a \( P_{5,\min} \) value of 0.116. So the joint p-value is the probability that any one of the 6 ways could occur. This probability is 0.024 and is our joint p-value. The joint p-values for all the panels are in Table 4 and confirm that the unit root can be rejected, but only for the full panel of 23 exchange rates.

### Table 4: Bootstrapped CIPS joint exact p-values

<table>
<thead>
<tr>
<th>Exchange rates</th>
<th>Joint p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0.14</td>
</tr>
<tr>
<td>17</td>
<td>0.21</td>
</tr>
<tr>
<td>23</td>
<td><strong>0.024</strong></td>
</tr>
</tbody>
</table>

Note: P-value significance is highlighted as before. The values are from the null DGPs with drift. The 6- and 23-exchange rate panels have almost exactly the same p-values if drift is not included. The joint p-value for the 17-exchange rate panel without drift is 0.25.

\[21\] In contrast, the occasionally used Bonferroni adjustment essentially assumes the tests are independent. Power is degraded if the tests are correlated. Simes (1986) proposes a revision that performs better than the basic version with correlated tests. Our approach, however, continues our strategy of using bootstrap techniques in order to base inference on characteristics of our specific data set.
A variation on this joint approach is to compute the $P_{\beta}$ value and corresponding $P_{x,\text{min}}$ values that give a joint p-value of 0.05. If any of the actual $P_x$ values are less than this $P_{x,\text{min}}$ value, then the unit root is rejected at the 0.05 level. The set of $P_{x,\text{min}}$ values can be considered 0.05 critical values for the joint test. For the 23 exchange rate panel (with drift in the DGP), the $P_{x,\text{min}}$ values are 0.003, 0.017, 0.045, 0.089, 0.166, and 0.313, each of which has a $P_{\beta}$ value of 0.0175. Two of the observed p-values are less than their counterpart on the $P_{x,\text{min}}$ list, supporting rejection at the 0.05 level.

6. Are the time orders of any trends in the stationary real exchange rates nonlinear?

We now turn to the question of whether, given the unit root rejection, nonlinear deterministic trends could be present in real exchange rates. It has already been noted that the time-order 2 CIPS test gives the strongest rejection, indirectly suggesting that quadratic trends are present. Now, additional evidence is presented.

Our first approach is to look at the real exchange rates individually, testing for the statistical significance of deterministic terms in autoregressions. This is hampered by not knowing which are stationary, because the distributions of standard $t$-statistics on deterministic terms in such regressions are affected by this. However, the $G_A$ test for nonlinear deterministic terms in autoregressions by Park and Choi (1988) avoids the problem; it is valid for both deterministic-trend-stationary and stochastic trend processes. According to bootstrapped distributions for the joint test of time order coefficients 2 through 5, the result is that no test statistic is significant for any real exchange rate at the 0.05 or 0.10 level.

This failure to find nonlinear trends could reflect low power. Therefore, we next try a panel approach to gain power. In order to do so, we assume that all the real exchange rates are stationary. Given this, standard chi square Wald tests on the coefficients of the deterministic terms in autoregressions will be valid. We compute the Wald tests from the set of 23 ADF real exchange rate regressions with 5th order time polynomials, where the SUR technique is used to

---

22 The unit root is rejected, but which of the real exchange rates are stationary? The CIPS procedure is not designed to answer this question. The recent approach of Pesaran, Smith, Yamagata, and Hvozdyk (forthcoming) could be applied. They estimate the proportion of all possible bilateral real exchange rates that are stationary among 50 countries, 1957-2001. We leave approaches such as this for future work.

23 The more recent tests of Vogelsang (1998), Bunzel and Vogelsang (2005), and Harvey, Leybourne, and Taylor (2007) could in principle be used to test for nonlinear trends in the face of the unknown order of integration. However, for none of these have certain specifics required to actually apply the tests to nonlinear trends (as opposed to linear trends) yet been computed.

24 To save space, we omit the detailed results. The failure to reject at the 0.05 level also follows from the asymptotically applicable chi square distribution, although there are few rejections at the 0.10 level. Regarding the bootstrap, because the test is valid with both unit root and stationary processes, we use both unit root DGPs (with constant drift) and linear trend DGPs estimated from the data. The unit root DGPs are the same as used for the CIPS test. The linear trend stationary DGPs are implemented analogously and with the same lag orders.

25 Note that this assumption is not as strong as the one of equal adjustment coefficients under the alternative hypotheses found in a number of the papers surveyed here, e.g., Papell and Theodoridis (2001).
gain efficiency. The lag orders are determined in a test-down for each regression. Thus, the equations are

\[
\Delta y_t = \sum_{q=0}^{5} a_q t^q a^2_{0,i} + b_i y_{t-1} + \sum_{j=1}^{9} \delta_{q} \Delta y_{t-j} + e_{it}
\]

(6)

for \( i = 1-23 \). Table 5 reports the bootstrapped p-values for a variety of individual and joint tests on the nonlinear terms for the system of 23 real exchange rates.

<table>
<thead>
<tr>
<th>Time orders</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>2-3</th>
<th>2-4</th>
<th>2-5</th>
<th>3-4</th>
<th>3-5</th>
<th>4-5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.20</td>
<td>0.30</td>
<td>0.41</td>
<td>0.51</td>
<td>0.050</td>
<td>0.037</td>
<td>0.060</td>
<td>0.050</td>
<td>0.076</td>
<td>0.069</td>
</tr>
</tbody>
</table>

No set of coefficients for any one time order is significant across the 23 real exchange rates; i.e., \( a_{2,i} = 0 \) for all 23 cannot be rejected, \( a_{3,i} = 0 \) for all 23 cannot be rejected, etc. However, several combinations of time order coefficients are statistically significant. Most noticeably, the null of \( a_{2,i} = a_{3,i} = a_{4,i} = 0 \) for all 23 can be rejected with a p-value clearly below 0.05. In combination with the failure to reject \( a_{5,i} = 0 \), it appears that if the nonlinear trends are indeed present, a time order as high as 5 is not needed to approximate them.

We should, however, once again consider the effects on significance levels of conducting multiple tests. Table 5 involves ten tests of the null of linear-trend stationarity against various forms of nonlinear trends, and so individual rejections could have occurred under the null with a higher probability than the nominal value of 0.05. Thus, we have computed a joint p-value for the ten tests in the same manner as for the six CIPS tests. The resulting p-value is 0.088. Consequently, we can conclude at the 0.10 level but not at the 0.05 level that nonlinear trends are present.

7. CIPS test behavior under mean, linear-trend, and nonlinear-trend stationarity

We now return to the CIPS test to examine its behavior under mean, linear, and nonlinear-trend stationarity to see what further light can be shed on the question of unit roots versus stationarity, particularly the possibility of stationarity around nonlinear trends. To begin with, we investigate the power of the tests. This is also motivated by the desire to explain the

\[\text{gain efficiency.}^{26}\]

Alternatively, the CADF equations could be used. However, they involve the estimation of coefficients for the cross-section means and their lags, which would decrease efficiency, lessening the ability to detect the nonlinear trends.

\[\text{The lag orders are determined in a test-down for each regression. Thus, the equations are}^{27}\]

The DGPs for the null hypothesis are linear-trend stationary, the same ones used to bootstrap the \( G_A \) test. There are 10,000 replications.

\[\text{The seeming contradiction between the individual and joint time term results could reflect a multicollinearity problem: the correlations among}^{28}\]

\( t^2, t^3, t^4, \text{ and } t^5 \) are very high: 0.932 to 0.995.
failure to reject with the smaller panels and higher time orders, which could reflect low power in these circumstances.

To confirm this possibility, we now present 0.05 size-adjusted powers for the individual CIPS tests. The bootstrapping proceeds as earlier, except the data-based DGPs are now stationary processes around trends of various time orders. With so many exchange rates and time orders, the potential number of combinations is rather large, so we only address cases where each exchange rate has the same time order as the others in the DGP. For lag orders, the same are used as before. To save space, we only report results where there is drift in the null distribution that is used to compute 0.05 critical values. See Tables 6a, 6b, and 6c. Results without drift in the null DGP are very similar. There are 10,000 replications of each deterministic-trend DGP.

**Table 6a: 0.05 Size-Adjusted Powers for the Individual CIPS Tests (6 exchange rates):**

<table>
<thead>
<tr>
<th>DGP time order</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.423</td>
<td>0.175</td>
<td>0.095</td>
<td>0.064</td>
<td>0.047</td>
<td>0.041</td>
</tr>
<tr>
<td>1</td>
<td>0.217</td>
<td>0.226</td>
<td>0.122</td>
<td>0.067</td>
<td>0.054</td>
<td>0.036</td>
</tr>
<tr>
<td>2</td>
<td>0.280</td>
<td>0.194</td>
<td>0.180</td>
<td>0.103</td>
<td>0.072</td>
<td>0.051</td>
</tr>
<tr>
<td>3</td>
<td>0.003</td>
<td>0.000</td>
<td>0.185</td>
<td>0.123</td>
<td>0.087</td>
<td>0.063</td>
</tr>
<tr>
<td>4</td>
<td>0.001</td>
<td>0.040</td>
<td>0.173</td>
<td>0.105</td>
<td>0.125</td>
<td>0.086</td>
</tr>
<tr>
<td>5</td>
<td>0.000</td>
<td>0.000</td>
<td>0.520</td>
<td>0.126</td>
<td>0.157</td>
<td>0.160</td>
</tr>
</tbody>
</table>

**Table 6b: 0.05 Size-Adjusted Powers for the Individual CIPS Tests (17 exchange rates):**

<table>
<thead>
<tr>
<th>DGP time order</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.856</td>
<td>0.398</td>
<td>0.187</td>
<td>0.115</td>
<td>0.086</td>
<td>0.065</td>
</tr>
<tr>
<td>1</td>
<td>0.518</td>
<td>0.577</td>
<td>0.294</td>
<td>0.164</td>
<td>0.112</td>
<td>0.081</td>
</tr>
<tr>
<td>2</td>
<td>0.307</td>
<td>0.240</td>
<td>0.435</td>
<td>0.257</td>
<td>0.168</td>
<td>0.118</td>
</tr>
<tr>
<td>3</td>
<td>0.000</td>
<td>0.000</td>
<td>0.279</td>
<td>0.323</td>
<td>0.204</td>
<td>0.138</td>
</tr>
<tr>
<td>4</td>
<td>0.000</td>
<td>0.000</td>
<td>0.119</td>
<td>0.051</td>
<td>0.303</td>
<td>0.186</td>
</tr>
<tr>
<td>5</td>
<td>0.000</td>
<td>0.000</td>
<td>0.027</td>
<td>0.811</td>
<td>0.614</td>
<td>0.402</td>
</tr>
</tbody>
</table>
Table 6c: 0.05 Size-Adjusted Powers for the Individual CIPS Tests (23 exchange rates):

<table>
<thead>
<tr>
<th>DGP time order</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.902</td>
<td>0.437</td>
<td>0.193</td>
<td>0.114</td>
<td>0.075</td>
<td>0.057</td>
</tr>
<tr>
<td>1</td>
<td>0.570</td>
<td>0.626</td>
<td>0.314</td>
<td>0.165</td>
<td>0.100</td>
<td>0.061</td>
</tr>
<tr>
<td>2</td>
<td>0.506</td>
<td>0.360</td>
<td>0.471</td>
<td>0.268</td>
<td>0.162</td>
<td>0.101</td>
</tr>
<tr>
<td>3</td>
<td>0.000</td>
<td>0.000</td>
<td>0.370</td>
<td>0.360</td>
<td>0.217</td>
<td>0.129</td>
</tr>
<tr>
<td>4</td>
<td>0.000</td>
<td>0.000</td>
<td>0.196</td>
<td>0.101</td>
<td>0.340</td>
<td>0.190</td>
</tr>
<tr>
<td>5</td>
<td>0.000</td>
<td>0.000</td>
<td>0.144</td>
<td>0.902</td>
<td>0.570</td>
<td>0.428</td>
</tr>
</tbody>
</table>

Within each panel, we find the expected trait that tests with both over- and under-specified time orders tend to lack power. The only exceptions are the ability of 3<sup>rd</sup> and 4<sup>th</sup> time-ordered tests to detect a 5<sup>th</sup> order nonlinear trend, which probably reflects some characteristics specific to our particular data. Regarding the primary question, the panel tests with fewer individual exchange rates do have lower power, as expected from Pesaran’s (2007) discussion. Therefore, the failure to reject with the smaller panels does not necessarily contradict the rejections with the largest panel.

The results also give clues about the most likely time order for the trend. If the nonlinear trend order were as high as 5, the 3<sup>rd</sup>, 4<sup>th</sup>, and 5<sup>th</sup> order CIPS tests would have a very good chance of rejecting the unit root, but there are no such rejections using the actual data. Likewise, if the trends’ time orders were 3 or higher, the 1<sup>st</sup> order (linear-trend) CIPS test would be extremely unlikely to give a rejection, yet it does using the actual data. On the other hand, if there were no trends, only constants (time order 0), then the 2<sup>nd</sup> order CIPS test would be unlikely to generate a rejection using the actual data, but it does. And the time-order-0 CIPS test should have given a strong rejection using the actual data, but it does not. These observations suggest linear or quadratic trends as the most likely.

The 0.05 size-adjusted power of the joint test of the CIPS tests for all 6 time orders in each panel can also be computed. This is done using the various critical values that yield joint 0.05 p-values under the null, as discussed at the end of section 5. The powers are in Table 7 for the case that includes drift in the unit root DGP used to get the critical values. As with the individual tests, the smaller panels suffer a power loss compared with the full panel. (And as with the individual CIPS tests, joint results using driftless unit root DGPs are very similar.)
Table 7: 0.05 Size-Adjusted Powers for the Joint Tests

<table>
<thead>
<tr>
<th>DGP time order</th>
<th>Exchange rates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td>0</td>
<td>0.180</td>
</tr>
<tr>
<td>1</td>
<td>0.155</td>
</tr>
<tr>
<td>2</td>
<td>0.200</td>
</tr>
<tr>
<td>3</td>
<td>0.065</td>
</tr>
<tr>
<td>4</td>
<td>0.090</td>
</tr>
<tr>
<td>5</td>
<td>0.257</td>
</tr>
</tbody>
</table>

As an alternative, we can compute “p-values” under the stationary processes with the various time trend orders. Such a p-value is the proportion of simulated CIPS test statistics for a given DGP that is less than the actual CIPS value. A p-value far from 0.50 means that the actual test statistic is far from the middle of the sampling distribution for the given trend order and thus not very consistent with that trend order. A p-value close to 0.50 means the test statistic is consistent with the given sampling distribution. The p-values are in Table 8 for the 23-exchange-rate panel (because it is the one that gives the joint rejection).

Table 8: P-values under stationarity

<table>
<thead>
<tr>
<th>DGP time order</th>
<th>Test time order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.923</td>
</tr>
<tr>
<td>1</td>
<td>0.622</td>
</tr>
<tr>
<td>2</td>
<td>0.559</td>
</tr>
<tr>
<td>3</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>0.000</td>
</tr>
</tbody>
</table>

According to these values, a time trend order of 3, 4, or 5 is extremely unlikely. That is because the CIPS test results specified for time orders 0 and 1 are strikingly different from what the simulations suggest should occur under these trends. If the six tests were viewed as “hypothesis tests” of these time orders, the hypotheses would be resoundingly rejected by the first two tests. The CIPS test specified with time order 3 also generates a “rejection” for time order 5, this time in the upper tail. On the other hand, the CIPS test specified with a constant is most consistent with a quadratic trend, as is the CIPS test with time order 2.

To get a different view of the p-value evidence, Table 9 shows the mean absolute deviation, mean squared deviation, and median absolute deviation of the p-values from 0.50 of the six CIPS tests under each DGP time order. For a point of reference, these measures are also
shown for the case of the unit root with drift. By all three measures, the 6 CIPS test results are overall most similar to those expected when the 23 real exchange rates have quadratic trends.

<table>
<thead>
<tr>
<th>DGP time order</th>
<th>Mean Abs. Dev.</th>
<th>Mean Sqd. Dev.</th>
<th>Median Abs Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>unit root w/drift</td>
<td>0.395</td>
<td>0.401</td>
<td>0.430</td>
</tr>
<tr>
<td>0</td>
<td>0.267</td>
<td>0.304</td>
<td>0.332</td>
</tr>
<tr>
<td>1</td>
<td>0.192</td>
<td>0.221</td>
<td>0.197</td>
</tr>
<tr>
<td>2</td>
<td>0.154</td>
<td>0.166</td>
<td>0.148</td>
</tr>
<tr>
<td>3</td>
<td>0.303</td>
<td>0.336</td>
<td>0.243</td>
</tr>
<tr>
<td>4</td>
<td>0.262</td>
<td>0.329</td>
<td>0.245</td>
</tr>
<tr>
<td>5</td>
<td>0.382</td>
<td>0.404</td>
<td>0.442</td>
</tr>
</tbody>
</table>

8. Summary and conclusion

We discuss the possibility of deterministic nonlinear trends rather than constants or linear trends as the alternative to unit roots in real exchange rates. To test for this, we extend the panel approach to testing for unit roots by adding nonlinear trends up to time order 5 to Pesaran’s (2007) CIPS test, which is then applied to 23 OECD-U.S. real exchange rates over the 1974-1998 post-Bretton Woods pre-euro era. We employ a conservative approach to statistical significance by using a bootstrapping procedure that accounts for many data-specific characteristics, data-based lag choice, and multiple applications of the test. The unit root can be rejected at the 0.05 level. This is consistent with previous unit root rejections in real exchange rate panels but follows from perhaps the most thorough bootstrapping approach yet. Moreover, across the CIPS tests specified with various time orders the strongest rejection occurs with quadratic trends, suggesting that the correct alternative is indeed nonlinear trends.

To further investigate this possibility, we compute the statistical significance of nonlinear trends in the real exchange rates. The results are somewhat inconclusive. While some tests are significant at the 0.05 level, the p-value that accounts for multiple testing only supports nonlinear trends at the 0.10 level. We then compare our specific CIPS test outcomes with simulated CIPS test behavior under various time trend orders. This indicates that our CIPS test outcomes are, overall, most consistent with CIPS test behavior under quadratic trends in the real exchange rates.

We thus conclude that the unit root should be rejected for at least some of our 23 real exchange rates and we believe the evidence points toward a nonlinear-trend-stationary process as the most likely alternative. If this is true, then PPP as commonly defined is violated, but real exchange rates do have meaningful long-run equilibria.
Appendix

Table A: The panel unit root literature on PPP

<table>
<thead>
<tr>
<th>Paper</th>
<th>Size</th>
<th>Base</th>
<th>Dates</th>
<th>Test</th>
<th>Adj spd</th>
<th>NL adj</th>
<th>Het ser</th>
<th>Cont corr</th>
<th>Dets</th>
<th>Data-based</th>
<th>Non-Norm</th>
<th>Cont corr</th>
<th>Ser corr</th>
<th>Lag sel</th>
<th>Rej</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abuaf-Jorion (1990)</td>
<td>10</td>
<td>US</td>
<td>73-87</td>
<td>GLS</td>
<td>=</td>
<td>No</td>
<td>No</td>
<td>GLS</td>
<td>M</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>a</td>
</tr>
<tr>
<td>Frankel-Rose (1996)</td>
<td>150</td>
<td>US</td>
<td>74-92</td>
<td>LL</td>
<td>=</td>
<td>No</td>
<td>?</td>
<td>C/TD</td>
<td>M</td>
<td>No</td>
<td>Yes</td>
<td>?</td>
<td></td>
<td></td>
<td>Yes</td>
<td>a, b</td>
</tr>
<tr>
<td>MacDonald (1996)</td>
<td>23</td>
<td>US</td>
<td>73-92</td>
<td>LL</td>
<td>=</td>
<td>No</td>
<td>Yes</td>
<td>None</td>
<td>M,L</td>
<td>No</td>
<td>Yes</td>
<td>c</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oh (1996)</td>
<td>111</td>
<td>US</td>
<td>73-89</td>
<td>LL</td>
<td>=</td>
<td>No</td>
<td>No</td>
<td>None</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>a, d</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wu (1996)</td>
<td>18</td>
<td>US</td>
<td>74-93</td>
<td>LL</td>
<td>=</td>
<td>No</td>
<td>No</td>
<td>C/TD</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jorion-Sweeney (1996)</td>
<td>10</td>
<td>US,G</td>
<td>73-93</td>
<td>GLS</td>
<td>=</td>
<td>No</td>
<td>No</td>
<td>GLS</td>
<td>M,L,B</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>e, f</td>
</tr>
<tr>
<td>Lothian (1997)</td>
<td>22</td>
<td>US</td>
<td>74-90</td>
<td>LL</td>
<td>=</td>
<td>No</td>
<td>No</td>
<td>None</td>
<td>M</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papell (1997)</td>
<td>20</td>
<td>US,G</td>
<td>73-94</td>
<td>GLS</td>
<td>=</td>
<td>No</td>
<td>Yes</td>
<td>GLS</td>
<td>M</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Usually</td>
<td>g, h</td>
</tr>
<tr>
<td>O'Connell (1998)</td>
<td>63</td>
<td>UK</td>
<td>73-95</td>
<td>GLS</td>
<td>=</td>
<td>No</td>
<td>No</td>
<td>GLS</td>
<td>M</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>i, j</td>
</tr>
<tr>
<td>Papell-Theodoridis (1998)</td>
<td>20</td>
<td>US,G</td>
<td>73-96</td>
<td>GLS</td>
<td>≠</td>
<td>No</td>
<td>Yes</td>
<td>GLS</td>
<td>M</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No, Yes</td>
<td>k</td>
</tr>
<tr>
<td>Taylor-Sarno (1998)</td>
<td>4</td>
<td>US</td>
<td>73-96</td>
<td>GLS</td>
<td>≠</td>
<td>No</td>
<td>Yes</td>
<td>GLS</td>
<td>M</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Papell-Theodoridis (2001)</td>
<td>20</td>
<td>many</td>
<td>73-96</td>
<td>GLS</td>
<td>=</td>
<td>No</td>
<td>Yes</td>
<td>GLS</td>
<td>M</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Usually</td>
</tr>
<tr>
<td>Taylor et al. (2001)</td>
<td>4</td>
<td>US</td>
<td>73-96</td>
<td>GLS</td>
<td>≠</td>
<td>Yes</td>
<td>No</td>
<td>GLS</td>
<td>M,L</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>m, n</td>
</tr>
<tr>
<td>Wu-Wu (2001)</td>
<td>20</td>
<td>US</td>
<td>73-97</td>
<td>IPS/MW</td>
<td>≠</td>
<td>No</td>
<td>Yes</td>
<td>None</td>
<td>M</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>?</td>
<td>Yes</td>
<td>o</td>
</tr>
<tr>
<td>Choi (2001)</td>
<td>6</td>
<td>US</td>
<td>73-96</td>
<td>IPS/Ch</td>
<td>≠</td>
<td>No</td>
<td>Yes</td>
<td>None</td>
<td>M</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smith et al. (2004)</td>
<td>17</td>
<td>US</td>
<td>88-98</td>
<td>IPS-p</td>
<td>≠</td>
<td>No</td>
<td>Yes</td>
<td>None</td>
<td>M</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>q</td>
</tr>
<tr>
<td>Papell (2006)</td>
<td>20</td>
<td>US</td>
<td>73-98</td>
<td>GLS</td>
<td>=</td>
<td>No</td>
<td>Yes</td>
<td>GLS</td>
<td>M</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Lopez-Papell (2007)</td>
<td>22</td>
<td>many</td>
<td>73-01</td>
<td>GLS</td>
<td>=</td>
<td>No</td>
<td>Yes</td>
<td>GLS</td>
<td>M</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Some</td>
<td>r</td>
</tr>
<tr>
<td>Pesaran (2007)</td>
<td>17</td>
<td>US</td>
<td>73-98</td>
<td>CIPS</td>
<td>≠</td>
<td>No</td>
<td>Yes</td>
<td>CIPS</td>
<td>M</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Lopez (2008)</td>
<td>20</td>
<td>several</td>
<td>73-01</td>
<td>GLS-p</td>
<td>=</td>
<td>No</td>
<td>Yes</td>
<td>GLS-p</td>
<td>M</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>s</td>
</tr>
<tr>
<td>Cerrato et al. (2008)</td>
<td>20</td>
<td>US</td>
<td>73-98</td>
<td>CIPS</td>
<td>≠</td>
<td>Yes</td>
<td>Yes</td>
<td>CIPS</td>
<td>M</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

Notes: See next page.
Notes to Table A:

Size: Number of exchange rates.
Base: Country used as numeraire.
Dates: Estimation period, rounded off to years.
Test: GLS = generalized least squares; LL – Levin and Lin (1993); IPS = Im, Pesaran, Shin (1997, 2003); MW = Maddala and Wu (1999); Ch = Choi (2001); IPS-p = a more powerful IPS from using the WS test (Pantula, Gonzalez-Farias, Fuller, 1994) and the Max test of Leybourne (1995) instead of the standard ADF; GLS-p = a more powerful GLS by applying SUR to the univariate DF-GLS test (Elliott, Rothenberg, Stock, 1996) instead of to the standard ADF test, as proposed by Lopez (forthcoming); CIPS = cross-section augmented IPS test of Pesaran (2007).
Adj spd: Is the adjustment speed under the alternative hypothesis equal or not across real exchange rates in the test specification?
NL adj: Is the adjustment to the long-run equilibrium nonlinear under the alternative hypothesis?
Het ser: Does the test specify heterogeneous serial correlation processes across the real exchange rates?
Cont corr: How, if at all, does the test itself (not the bootstrap, if present) control for contemporaneous (cross-section) correlation?
GLS = generalized least squares/SUR; C/TD = cross-section demeaning or time dummies (not generally effective, see, e.g., Pesaran, 2007); CIPS = Pesaran’s (2007) approach.
Dets: What deterministic terms are specified? M = means; L = linear trends; B = breaks in means.
Data-based: Is the sampling distribution for getting p-values or critical values data based? (Only adjusting for actual sample size is not counted.)
Non-norm: Do the residuals used to build up simulated data sets incorporate non-normality in the actual data?
Cont corr: Do the residuals used to build up simulated data sets incorporate contemporaneous correlation in the actual data?
Ser corr: In the data-generating process, are data-based serial correlation processes included?
Lag sel: Is the same lag selection process that was used with the actual data also applied to each simulated data set?
Rej: Our interpretation of the paper’s general conclusion on rejecting unit roots.
Notes: a Additional, longer time periods are in the paper; the rejection conclusion we report refers to the period listed under “Dates.”
b The rejection applies only to the case of homogeneous constants across exchange rates and no time dummies.
c Rejection occurred both without and with linear trends.
d Two subsets were also tested; the rejection conclusion also applies to the smaller 23 OECD case.
e According to the paper's notations, the tests specified no serial correlation at all.
f The rejections occurred with means but not with linear trends or a break in mean.
g The rejections occurred for both quarterly and monthly data with Germany, but only monthly for the U.S.
(continued next page)
There are rejections in some but not all of several sub-panels.

O'Connell's estimation procedure makes the choice of base country irrelevant.

None of four sub-panels rejected, either.

Not rejected with the U.S. as base; rejected with Germany as base.

The test rejects with most countries as the base, but not with the U.S.

There was no serial correlation correction because Taylor et al. concluded that first order systems in levels were adequate with unit roots imposed.

Unit roots were rejected with linear trends specified as well as without; Taylor et al. (2001) concluded the trends were insignificant.

Several sub-panels did not reject.

The conclusion to reject is firm only if Japan is excluded from the panel.

Smith et al. (2004) contained a significant data error; the rejection conclusion follows from the corrected results in Pesaran (2007).

All tests are applied to geographical sub-panels of the 22 real exchange rates, with numerous alternative base countries for each.

With the US as base, the rejections are found for most of several sub-panels as well as for the full panel; with other counties as base, rejections are less frequent.
References


Mundell, R., 1962. The appropriate use of monetary and fiscal policy for internal and external stability. IMF Staff Papers 9, 70-77.


