ELSEVIER

Journal of International Money and Finance
17 (1998) 597–614

Journal of International Money and Finance

Parity reversion in real exchange rates during the post-Bretton Woods period

Yin-Wong Cheung[a], Kon S. Lai[b,*]

[a]*University of California, Santa Cruz, CA 95064, USA*
[b]*California State University, Department of Economics, Los Angeles, CA 90032, USA*

Abstract

A common view among recent studies on purchasing power parity is that the post-Bretton Woods period is far too short to reveal any significant parity reversion in individual series of real exchange rates. Is this really so? The answer, this study shows, depends very much on the statistical test being used. Two efficient univariate unit-root tests are applied to uncover parity reversion. These tests require much shorter sample sizes than conventional tests to attain the same statistical power. Empirical results show that parity reversion can be unveiled over the modern float if an efficient unit-root test is applied.

JEL classification: F31, F41

Keywords: Real exchange rate; Purchasing power parity; Efficient unit-root test; Test power

1. Introduction

The relevance of purchasing power parity (PPP) has been a hotly debated issue, witness the extensive survey studies by Breuer (1994), Froot and Rogoff (1995) and Rogoff (1996). Prior studies on long-run PPP typically find less than favorable evidence in the post-1973 data, suggesting that PPP deviations are governed by permanent disturbances during the post-Bretton Woods period, unlike other historical periods. The failure to find support for long-run PPP during the current

*Corresponding author. Tel.: +1 213 3432930; fax: +1 213 3435462; e-mail: klai@calstatela.edu

0261-5606/98/$19.00
PII: S0261-5606(98)00020-5

float highlights a basic testing problem. PPP deviations can be slow to reverse, and conventional econometric techniques have low power to identify stationary but persistent dynamics.

An approach to addressing the low-power problem is to expand the sample period. The failure to detect parity reversion has often been attributed to the short span of the post-Bretton Woods period. Many recent studies explore data from other historical periods of exchange rate arrangements, dating sometimes back to the gold standard (Abuaf and Jorion, 1990; Ardeni and Lubian, 1991; Diebold et al., 1991; Glen, 1992; Cheung and Lai, 1993a; Culver and Papell, 1995). Early testing of long-run PPP using long historical data is conducted by Frankel (1986), Edison (1987) and Edison and Klovland (1987). In most cases, long-sample data are obtained by combining fixed-rate and floating-rate regimes together. Empirical results based on these data mostly show support for long-run PPP.

The issue still not definitely resolved is whether PPP has collapsed since the advent of the modern float. The long-sample evidence for long-run PPP documented so far has come from data sets in which the recent float data are at best a small proportion only. Many of the foregoing studies note that PPP reversion cannot be found for the post-Bretton Woods portion of their data. On the other hand, parity reversion has easily been found in earlier studies for other historical periods (Enders, 1988; Taylor and McMahon, 1988). Systematic differences in the behavior of real exchange rates under different exchange rate arrangements have also been noted by Mussa (1986) and Baxter and Stockman (1989), among others. It is therefore not clear whether the findings based on long-sample data confirm simply the presence of parity reversion in the pre-modern float period or show its presence over the modern float as well. Lothian and Taylor's (1996) study represents the first effort to offer some indirect evidence supporting the latter. In examining the dollar/pound and franc/pound real rates, these authors observe no significant evidence of a structural change between the pre- and post-Bretton Woods periods.

The analysis by Lothian and Taylor (1996) is important. The basic motivation behind the long-sample approach is that the post-Bretton Woods period is far too short to reveal PPP reversion, thereby leading to the widespread failure to detect it. If these authors' results can be shown to be generally applicable to other long-sample series of real exchange rates, they will provide a needed justification for using data information from other historical periods to indirectly infer the behavior of real exchange rates during the current float.

This study takes a more direct approach and provides considerable evidence of parity reversion from the recent float data directly. The findings can reinforce those from Lothian and Taylor (1996) and further strengthen the empirical support for parity reversion. Without using data from different historical periods to enlarge the sample size, an alternative approach to handling the power deficiency problem is to apply efficient tests with improved power. Statistical procedures can differ substantially in terms of efficiency and test power. A more efficient test may need much less observations to achieve a specific level of power than a less efficient test. Accordingly, using tests with improved efficiency has effectively the same implication for power as having more observations available, while abstracting from the potential issue concerning cross-sample heterogeneity in mixing data from different time periods. In this regard, recent advances in econometric analysis in providing increasingly powerful tests may prove useful.

Cheung and Lai (1993b) seek to fill the gap in the empirical support for long-run PPP by examining a weaker version of PPP, which requires only that deviations from a linear combination of exchange rates and national price levels be stationary (see also Edison et al., 1997). Using multivariate cointegration tests, significant evidence of a long-run relationship is found between exchange rates and price levels over the recent float. Froot and Rogoff (1995) caution, however, that the interpretation of the cointegration results may not be clear. Due to serious small-sample bias, the long-run coefficient estimates obtained for the recent floating-rate data can vary widely across country pairs, making them difficult to interpret unless measurement errors in prices are very substantial. Consequently, Froot and Rogoff (1995) favor the use of unit-root tests, which analyze directly whether the real exchange rate is mean-reverting. The major problem remains: commonly used unit-root tests lack sufficient power to uncover parity reversion in real exchange rates. To be sure, there are theoretical reasons to suggest that PPP reversion, if indeed it exists, can be sluggish. For example, intertemporal smoothing of traded goods consumption (Rogoff, 1992) or cross-country wealth redistribution effects (Obstfeld and Rogoff, 1995) may generate highly persistent dynamics for the real exchange rate. As a result, the possible use of unit-root tests that have good power against persistent alternatives should clearly be desirable.

This study re-examines the validity of parity reversion in real exchange rates during the post-Bretton Woods period by making use of some recent developments in unit-root testing. In studying the asymptotic power envelope for various unit-root tests, Elliott et al. (1996) propose a simple modification of the augmented Dickey–Fuller (ADF) test such that the modified test can nearly achieve the power envelope using generalized least squares (GLS) estimation. The resulted DF-GLS test is shown to be approximately uniformly most powerful. Monte Carlo results confirm that the power improvement from using the DF-GLS test can be large relative to the standard ADF test. Park and Fuller (1995), on the other hand, recommend the use of a modified Dickey–Fuller test based on weighted symmetric least squares (WSLS) estimation. Pantula et al. (1994) show that this test, hereafter called the DF-WS test, has similar power as the DF-GLS test. These two efficient tests are both used to uncover PPP reversion in the post-1973 data, and they require much shorter sample sizes than conventional tests to attain the same test power (Stock, 1994). The efficiency gains are shown to be critically important for a proper evaluation of the stationarity property of real exchange rates. In contrast to usual unit-root tests, the efficient tests are found to yield considerably more evidence in favor of no unit root in real exchange rates.

The use of efficient unit-root tests to uncover PPP reversion shares a similar spirit with the multivariate unit-root analysis in recent panel studies of PPP (Frankel and Rose, 1996; Oh, 1996; Lothian, 1997; Papell, 1997). Motivated by the work of Levin and Lin (1992), these panel studies consider pooling data across currencies to raise statistical power, without extending the time span of sample data. Efficiency gains are obtained for unit-root tests by exploiting the additional data variation in panel data and imposing some cross-equation restrictions. With the help of improved test efficiency, panel analyses generally succeed in uncovering evidence in favor of PPP reversion for the post-war or current float period. Papell (1997) cautions, nonetheless, that the inferences based on the panel method can be sensitive to sample selection, in particular, to the size of the panel as well as the grouping of countries. O'Connell (1996) also illustrates that panel unit-root tests can produce biased results in the presence of cross-sectional dependence. Unlike panel tests, the efficient unit-root tests applied in this study remain univariate procedures, with higher efficiency being attained by constructing the tests to be optimal against persistent local alternatives. Empirical results show that these univariate procedures can be useful additions to the menu of efficient tests for unveiling PPP reversion.

2. Purchasing power parity

The PPP hypothesis suggests the presence of a long-run equilibrium relationship between national price levels of two countries when expressed in common currency units. In allowing for short-run deviations, an empirical representation of the PPP relationship is given by

$$s_t = \gamma_0 + \gamma_1(p_t - p_t^*) + u_t, \tag{1}$$

where s_t is the logarithm of the spot exchange rate (domestic price of foreign currency); p_t and p_t^* are, respectively, the logarithms of the domestic and foreign price indexes; and u_t is an error term capturing deviations from PPP. An excellent review of the PPP theory and its different stages of tests has been provided by Froot and Rogoff (1995). Early empirical studies on PPP were interested mainly in the value of the slope coefficient, γ_1. Specifically, a simple test for PPP would entail an examination of whether the slope coefficient differs significantly from unity. Little attention was paid to the stochastic properties of the dynamics of adjustments toward PPP and their possible implications for statistical analysis, although significant disturbances to PPP were allowed for.

A major problem with the early studies on PPP was the failure to deal with the possible non-stationarity of exchange rates and prices. This problem prompted an alternative approach to testing PPP by evaluating the stationarity of the real exchange rate, y_t ($\equiv s_t - p_t + p_t^*$), with the coefficient $\gamma_1 = 1$ being imposed and not estimated. For the PPP to hold in the long-run, y_t should be stationary and not governed by permanent shocks. Nevertheless, the ability of this approach to uncover and confirm PPP reversion, even if it exists, has been called into question because of the unsatisfactorily low power of standard unit-root tests. Such shortcoming appears particularly severe in analyzing the limited span of data for the recent float.

Another approach considers tests of a weaker version of PPP obtained from relaxing the usual symmetry and proportionality restrictions. This approach examines the cointegration property between exchange rates and prices by examining whether some linear combination, $s_t - \gamma_1 p_t + \gamma_2 p_t^*$, is stationary. Although the cointegration approach seems able to yield evidence consistent with a weak form of PPP for the recent float data, its use has been criticized because of the lack of clear meaning of the cointegrating coefficient estimates, making the cointegration test results hard to interpret (Froot and Rogoff, 1995). Consequently, the present analysis will adopt the unit-root test approach but handle the power deficiency problem directly by using efficient, optimal tests.

3. Data and preliminary analysis

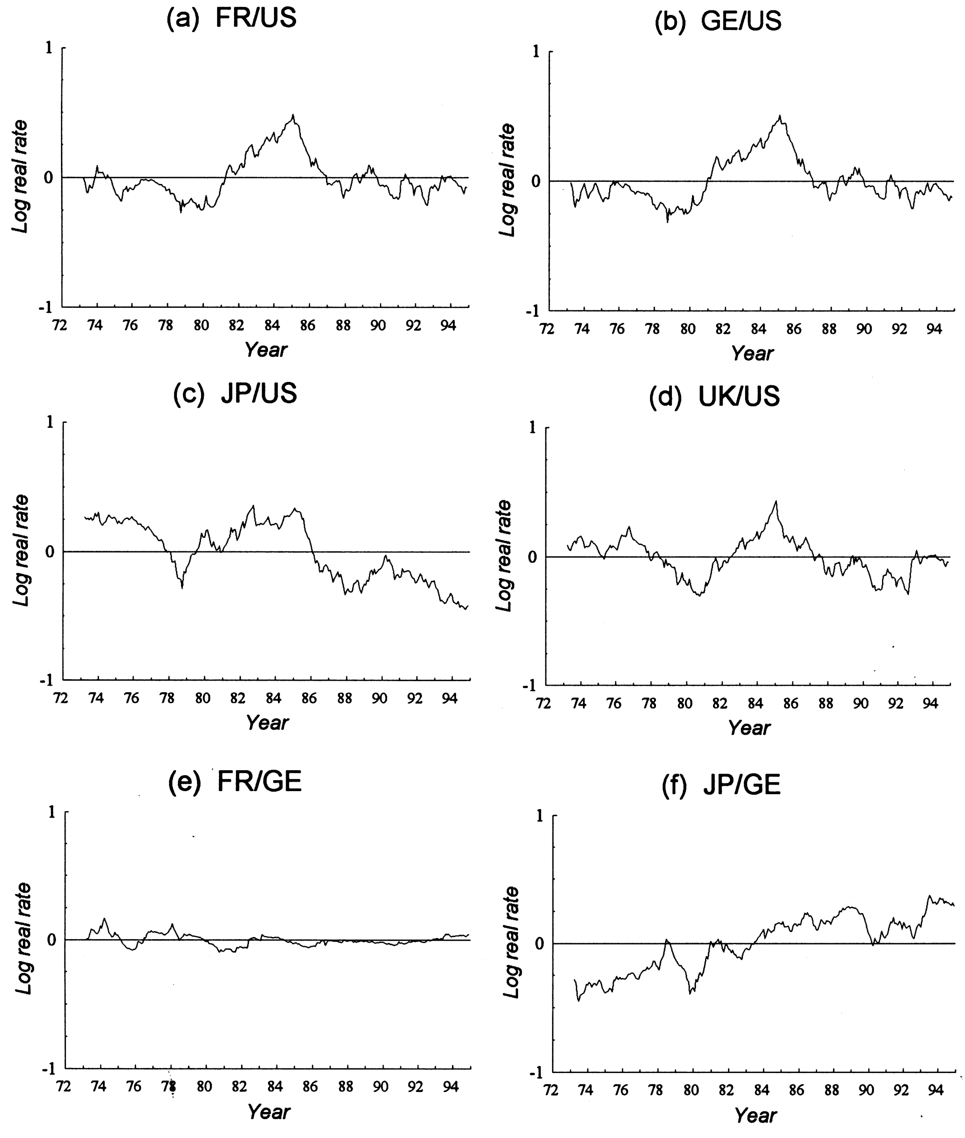
Real exchange rates during the recent floating-rate period for five countries — France (FR), Germany (GE), Japan (JP), the United Kingdom (UK) and the United States (US) — are studied. The data, taken from the International Monetary Fund's *International Financial Statistics* data tape, include month-end spot exchange rates (monthly averages yield similar results) and monthly price levels measured by consumer price indexes. All data series are seasonally unadjusted, covering the sample period from April 1973 through December 1994. Fig. 1 exhibits charts of real exchange rates for 10 possible bilateral cases. Each series is plotted and centered with respect to its sample mean. Despite the wide fluctuations in real exchange rates, some patterns of reversal can be observed in these charts. Nevertheless, the significance of the reverting tendency has been hard to establish statistically.

To serve as a benchmark for comparison, all series of real exchange rates are first tested for a unit root using the ADF test. For long-run PPP to hold, the real exchange rate should be stationary and contain no unit root. The ADF test involves estimating the following regression:

$$(1 - L)y_t = \mu_0 + \mu_1 t + \beta_0 y_{t-1} + \sum_{j=1}^{p} \beta_j (1 - L) y_{t-j} + \epsilon_t, \qquad (2)$$

where L is the lag operator and ϵ_t is the error term. The null hypothesis of a unit root is represented by $\beta_0 = 0$. The ADF statistic is given by the usual t-statistic for the β_0 coefficient.

Sequential unit-root tests devised by Banerjee et al. (1992), henceforth BLS, are also performed. These sequential tests extend the ADF test by accounting for possible trend shifts or breaks in the underlying data process. Culver and Papell (1995) illustrate that incorporating a trend break in the ADF test can help detect parity reversion better than the standard ADF test for real exchange rates under the gold standard. Perron and Vogelsang (1992) also provide evidence of parity reversion with broken trends for some long-horizon series of dollar-based real exchange rates. The trend-break hypothesis can be relevant in PPP analysis when substantial changes in differential productivity growth in tradables and non-trada-



bles occur across countries — the oft-called Balassa–Samuelson effect. To see whether the trend-break hypothesis is relevant for the post-Bretton Woods data, the BLS procedure is carried out here.

Consider the following representation of the $\{y_t\}$ process:

$$(1 - L)y_t = \mu_0 + \mu_1 t + \mu_2 d_t(k) + \beta_0 y_{t-1} + \sum_{j=1}^{p} \beta_j (1 - L) y_{t-j} + \zeta_t, \tag{3}$$

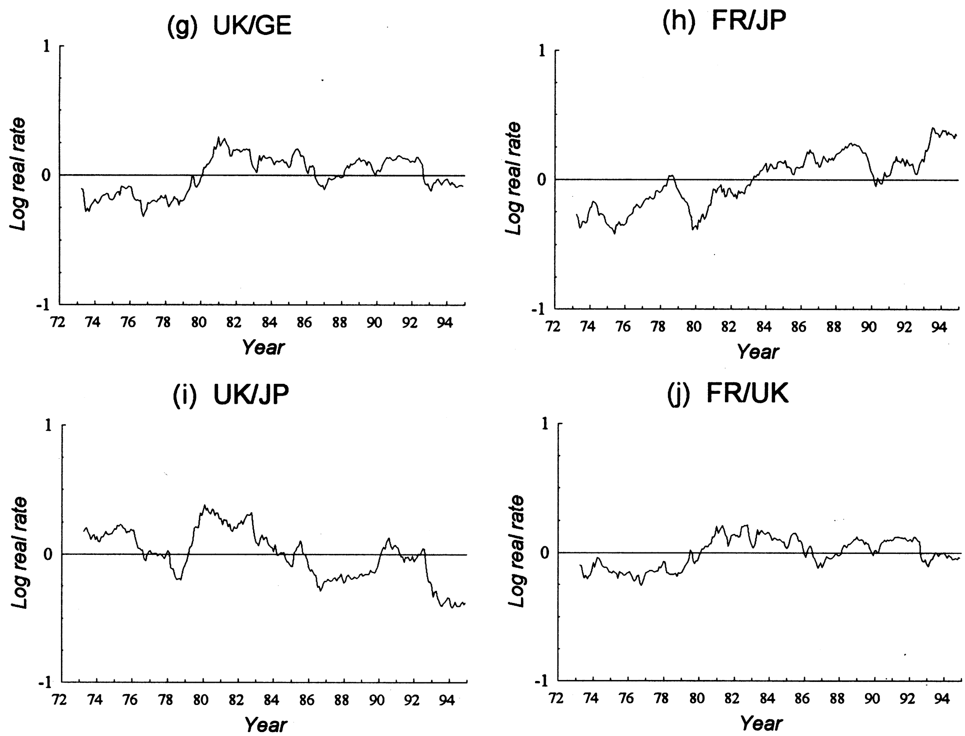


Fig. 1. Plots of the log of the real exchange rate.

where $d_t(k)$ is a dummy variable and ζ_t is the error term. When a trend shift is allowed for at time k, $d_t(k) = (t - k)I(t > k)$, with $I(\cdot)$ being the indicator function. Alternatively, when a mean shift (or a break in the trend) is allowed for at time k, $d_t(k) = I(t > k)$. For the usual ADF test, $d_t(k) = 0$. A sequence of statistics, $\tau_{DF}(k)$, indexed by k can be generated by varying k over the sample. BLS discuss several versions of the mean-shift or trend-shift sequential test. The minimal sequential test is employed in this study, and its test statistic is defined by

$$\tau_{DF}^{\min} \equiv \min_{r \leq k \leq T-r} \tau_{DF}(k) \tag{4}$$

for the sample size, T, and a trimming parameter, r. Following BLS, r is set equal to $[0.15T]$.

The results of the ADF test are summarized in Table 1. Both ADF tests with and without a time trend are conducted. For cases in which the time trend is significant at the 10% level, results of the ADF test with a time trend are reported. For cases of an insignificant time trend, however, results of the test without a time trend are reported. Evidently, the ADF test results are far from favorable to long-run PPP. Out of the 10 cases under consideration, in no more than two cases (GE/JP and

Table 1
Results from the ADF test

Country pair	Test type	p	ADF test Statistic
FR/US	No trend	2	−1.823
GE/US	No trend	2	−1.863
JP/US	With trend	6	−1.991
UK/US	No trend	2	−2.156
FR/GE	No trend	4	−2.818*
JP/GE	With trend	4	−3.418**
UK/GE	No trend	3	−2.375
FR/JP	With trend	2	−2.773
UK/JP	With trend	3	−2.294
FR/UK	No trend	3	−2.385

Notes. In cases in which tests with a time trend are used, the time trend variable is significant at the 10% level or better. The column beneath 'p' gives the lag used for the corresponding test (a maximum lag of $p = 10$ has been considered in the lag choice using the AIC together with residual analysis). Lag-adjusted finite sample critical values for the ADF test are based on Cheung and Lai (1995a). Statistical significance is indicated by a single asterisk (*) for the 10% level and a double asterisk (**) for the 5% level.

FR/GE) can significant evidence of stationarity be found. Hence, in accordance with previous findings, the ADF test reveals little evidence of parity reversion.

Table 2 contains the results of both mean-shift and trend-shift sequential tests. In no more than one case (GE/JP) can the null hypothesis of a unit root be rejected when a mean shift is included under the alternative hypothesis. For the trend-shift sequential test, the results are uniformly negative for parity reversion. It follows that even allowing for possible mean shifts or trend shifts in real exchange rates, no new empirical support for long-run PPP can be found. Hence, the trend-break hypothesis cannot explain the empirical failure to uncover PPP reversion during the current float.

4. Efficient unit-root tests

In testing for a unit root in the real exchange rate, the power of the statistical test used is of critical importance. The empirical failure to find reversion toward PPP can be resulted from deficiencies in test power. Two modified Dickey–Fuller tests with good power are applied in this study. Since both tests are relatively new, they are discussed below.

4.1. The DF-WS test

Park and Fuller (1995) devise a modified Dickey–Fuller test using the WSLS

Table 2
Results from two BLS sequential tests

Country pair	BLS sequential test			
	p	Mean-shift	p	Trend-shift
FR/US	2	−3.909	2	−2.401
GE/US	2	−4.077	2	−2.632
JP/US	3	−3.364	3	−2.621
UK/US	1	−3.303	1	−2.347
FR/GE	2	−3.943	2	−3.600
JP/GE	4	−4.958**	4	−3.918
UK/GE	3	−4.433	1	−3.250
FR/JP	4	−4.227	4	−3.603
UK/JP	4	−3.870	4	−2.958
FR/UK	1	−4.508	1	−3.936

Notes. The column beneath 'p' gives the lag used for the corresponding test (a maximum lag of $p = 10$ has been considered in the choice based on the AIC together with residual analysis). Critical values for the BLS sequential $\tau_{DF}^{\min}$ tests are provided by Banerjee et al. (1992; Table 2). For the sequential mean-shift $\tau_{DF}^{\min}$ test, the 10% and 5% critical values are, respectively, given by −4.51 and −4.80 for $T = 250$. For the sequential trend-shift $\tau_{DF}^{\min}$ test, the 10% and 5% critical values are given correspondingly by −4.12 and −4.39 for $T = 250$. Statistical significance is indicated by a double asterisk (**) for the 5% level.

estimator, which is more efficient than the OLS estimator in estimating autoregressive (AR) parameters. Consider the following specification of the $\{y_t\}$ process:

$$y_t = \rho y_{t-1} + \sum_{j=1}^{p} \alpha_j (1-L) y_{t-j} + e_t, \tag{5}$$

where ρ is the largest AR root and e_t is the error term. The DF-WS test involves minimizing a weighted sum of squared errors with respect to ρ and α:

$$\begin{aligned} Q(\rho,\alpha) = & \sum_{t=p+2}^{T} w_t \left[y_t - \rho y_{t-1} - \sum_{j=1}^{p} \alpha_j (1-L) y_{t-j} \right]^2 \\ & + \sum_{t=p+2}^{T} (1-w_{t-p}) \left[y_{t-p-1} - \rho y_{t-p} + \sum_{j=1}^{p} \alpha_j (1-L) y_{t-p+j} \right]^2, \end{aligned} \tag{6}$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_p)$ and the weight, w_t $(t = 1, 2, \ldots, T)$, is specified by

$$\begin{aligned} w_t &= 0 && \text{for} \quad 1 \le t \le p+1; \\ &= (t-p-1)/(T-2p) && \text{for} \quad p+1 < t \le T-p; \\ &= 1 && \text{for} \quad T-p < t \le T. \end{aligned} \tag{7}$$

Park and Fuller (1995) discuss the use of similar weighting schemes in various other studies to raise estimation efficiency by exploiting the time-reversibility property of stationary processes. With the weights and the minimization criterion, the initial and last observations are treated equally in the WSLS estimator. When $w_t = 1$ for all t, it reduces to the OLS case of the ADF test. The hypothesis $H_0:\rho = 1$ can be tested against $H_a:\rho < 1$ using the following statistic:

$$\tau_{ws} = \{V(\hat{\rho})\}^{-1/2}(\hat{\rho} - 1) \tag{8}$$

where $V(\hat{\rho})$ is the estimated variance of $\hat{\rho}$ obtained from the WSLS regression, with the error variance σ_e^2 being estimated by $Q(\hat{\rho},\hat{\alpha})/(T-2)$. The limiting distribution of τ_{ws} has been analyzed by Pantula et al. (1994) and Park and Fuller (1995).

To account for the unknown mean and possibly a deterministic time trend in the data process, $\{y_t\}$ can be demeaned and detrended prior to the WSLS regression for the DF-WS test. Specifically, y_t is first replaced with y_t^* before the WSLS regression, with y_t^* being constructed as: $y_t^* = y_t - z_t\gamma$, where $z_t = (1,t)$ and γ is the OLS estimator obtained from regressing y_t on z_t.

4.2. The DF-GLS test

Elliott et al. (1996), henceforth ERS, obtain the asymptotic power envelope for unit-root tests by analyzing the sequence of Neyman–Pearson tests of the null hypothesis $H_0:\rho = 1$ against the local alternative $H_a:\rho = 1 + \bar{c}/T$, where $\bar{c} < 0$. Based on asymptotic power calculation, ERS show that a modified Dickey–Fuller test, called the DF-GLS test, can achieve a substantial gain in power over traditional unit-root tests. The superior power performance of modified Dickey–Fuller tests has been documented by Pantula et al. (1994) and Stock (1994).

The DF-GLS$^\tau$ test that allows for a linear time trend is based on the following regression:

$$(1-L)y_t^\tau = \phi_0 y_{t-1}^\tau + \sum_{j=1}^{p} \phi_j(1-L)y_{t-j}^\tau + \nu_t, \tag{9}$$

where ν_t is an error term; and y_t^τ the locally detrended data process under the local alternative of $\bar{\rho} = 1 + \bar{c}/T$, is given by

$$y_t^\tau = y_t - z_t\beta, \tag{10}$$

with β being the least squares regression coefficient of $\tilde{y}_t$ on $\tilde{z}_t$, for which $\tilde{y}_t = [y_1, (1-\bar{\rho}L)y_2,..., (1-\bar{\rho}L)y_T]'$ and $\tilde{z}_t = [z_1, (1-\bar{\rho}L)z_2,..., (1-\bar{\rho}L)z_T]'$. The DF-GLS$^\tau$ statistic is given by the t-ratio, testing $H_0:\phi_0 = 0$ against $H_a:\phi_0 < 0$. ERS recommend that the parameter defining the local alternative, $\bar{c}$, be set equal to -13.5. For the test without a time trend, denoted by DF-GLS$^\mu$, it involves the same procedure as the DF-GLS$^{\tau}$ test, except that y_t^{τ} is replaced with the locally demeaned series y_t^{μ} and $z_t = 1$. In this case, the use of $\bar{c} = -7$ is recommended.

5. Empirical results

Results of the DF-WS and DF-GLS tests are summarized in Table 3. The lag parameters for both the DF-WS and DF-GLS tests are selected using the Akaike information criterion (AIC), as recommended by Pantula et al. (1994). In a number of cases, higher lag orders than those suggested by the AIC lag selection procedure are used to remove significant serial correlation in the residuals. As a result, our final lag choices are in effect the same as those using serial correlation as a selection criterion. On the other hand, although using the AIC alone would suggest smaller lags than those chosen based on serial correlation, the unit-root test results are qualitatively the same in terms of the number of rejection cases, independent of whether the AIC or serial correlation is used for lag selection.

Tests with and without a time trend are both conducted. For cases in which the time trend is statistically significant at the 10% level, results of tests with a time trend are reported. For cases of an insignificant time trend, however, results of tests without a time trend are reported. According to our test results, a significant time trend is found for all the series of real exchange rates involving the Japanese yen. The results are consistent with some previous findings presented by, e.g. Obstfeld (1993), who identified the presence of a deterministic time trend in,

Table 3
Results from modified Dickey–Fuller tests

Country pair	Test type	DF-WS		DF-GLS	
		p	Statistic	p	Statistic
FR/US	No trend	2	−2.081	2	−1.825*
GE/US	No trend	2	−2.071	2	−1.852*
JP/US	With trend	2	−2.158	3	−2.128
UK/US	No trend	2	−2.348*	2	−1.835*
FR/GE	No trend	4	−2.843**	4	−2.712**
JP/GE	With trend	3	−3.411**	3	−3.318**
UK/GE	No trend	3	−2.109	3	−1.779*
FR/JP	With trend	7	−3.690**	6	−3.416**
UK/JP	With trend	3	−2.529	3	−2.292
FR/UK	No trend	2	−2.264*	3	1.763*

Notes. In cases in which tests with a time trend are used, the time-trend variable is significant at the 10% level or better. The column beneath 'p' gives the lag used for the corresponding test (a maximum lag of $p = 10$ has been considered in the choice based on the AIC together with residual analysis). Lag-adjusted finite sample critical values for the DF-GLS test are based on Cheung and Lai (1995b). For the DF-WS test, lag-adjusted finite sample critical values are also estimated using response surface analysis (see Table 4). Statistical significance is indicated by a single asterisk (*) for the 10% level and a double asterisk (**) for the 5% level.

specifically, real exchange rates of the Yen. This author further demonstrated that such a time trend might be caused by differential productivity growth in tradables and non-tradables.

For the DF-WS test, as reported in Table 3, in five out of 10 cases (including the two bilateral rates involving British pounds examined in Lothian and Taylor (1996) long-sample study) can the hypothesis of a unit root be rejected in favor of stationary alternatives at the 10% significance level or better. These results, albeit somewhat mixed, are more favorable to the PPP hypothesis than the results of the ADF test.

When compared with the DF-WS test results, those from the DF-GLS test render much wider and stronger support for parity reversion in real exchange rates. The unit-root hypothesis can be rejected in favor of stationary alternatives for real exchange rates in eight out of the 10 cases, namely, FR/US, GE/US, UK/US, FR/GE, JP/GE, UK/GE, FR/JP and FR/UK. In testing for long-run PPP under the null hypothesis that exchange rates and prices are cointegrated rather than not cointegrated, Fisher and Park (1991) find greater difficulty in detecting PPP reversion in US-based data than German-based data. In contrast to the bivariate cointegration test used by Fisher and Park (1991), the DF-GLS test is a univariate test known to be optimal against persistent local alternatives. According to our results, stationarity in the real exchange rate can be uncovered in three out of the four US-based cases using an efficient unit-root test; whereas, stationarity can be detected in all the four Germany-based cases. Hence, more evidence in favor of PPP reversion can still be found when the German mark, rather than the US dollar, is used as the base currency. In this respect, our results accord with those of Fisher and Park (1991).

The use of efficient tests clearly produces more power in rejecting the unit-root hypothesis than the standard tests, yielding much evidence consistent with PPP reversion. Nonetheless, the rejections in some cases are at the 10% level of significance, which are weaker rejections than those at the 5% level. These weak rejections confirm the difficulty of unveiling mean reversion in data of short time spans. In this regard, some researchers (e.g. Johansen and Juselius, 1990) observe that a 5% test size may sometimes be too conservative for tests for non-stationarity when conducted on data of limited spans. It may be interesting to check if future research using additional data can yield more decisive rejections.

It should also be noted that among the countries considered in our study, France and Germany have been key participants in the European Monetary System's (EMS) quasi-fixed exchange rate mechanism, under which intra-EMS rates are allowed to move within a grid of bilateral fluctuation bands. These exchange rate bands had typically been narrow over time, though they were widen substantially under the pressure of speculative attacks in August 1993. The UK joined the exchange rate mechanism in October 1990 but then pulled out of it in September 1992. Since the inception of the EMS in March 1979, the EMS has gone through periodic currency realignments, most of which took place between 1979 and 1986. An issue concerns the potential effects of currency realignments on unit-root analysis. Under each realignment, the EMS rates would fluctuate within different bands. These shifts in stochastic processes may bias unit-root tests toward supporting the unit-root null too often. Such consideration should only strengthen our finding of stationarity, nonetheless, given that the unit-root hypothesis can be rejected in the FR/GE case here. In addition, the earlier reported results of the BLS sequential tests reveal no significant breaks in the corresponding series of real exchange rates. The evidence may reflect that France and Germany have spent special efforts on minimizing their relative movements in exchange rates.

Further insights into the unit-root test results can be gained by studying the persistence of the dynamics of real exchange rates. The value of the dominant AR root for each individual series is estimated. All the dominant AR roots are estimated to be less than unity, ranging from 0.934 to 0.974. The cases of FR/US and GE/US yield the two largest values of the AR root; whereas, the cases of FR/GE and JP/GE give the two smallest AR values. Indeed, US-based real exchange rates consistently have the dominant AR roots closer to unity than the other non-US-based rates, like those involving Germany. Persistence estimates are also obtained for individual cases using cumulative impulse response estimation (Campbell and Mankiw, 1987). These estimates again indicate that US-based real exchange rates generally display higher persistence than the others. Such difference in persistence can be responsible for the extra difficulty in detecting parity reversion in dollar-based real exchange rates, as observed elsewhere by Fisher and Park (1991). The relatively low persistence found in both FR/GE and JP/GE series may account for the ability of the ADF test to reject the unit-root hypothesis in these two cases, without using the more efficient DF-GLS test. For the other more persistent series, on the other hand, the failure of the ADF test to find parity reversion confirms its low power against stationary but persistent alternatives.

The differences in results between the ADF and DF-GLS tests highlight the critical importance of test power for properly evaluating the parity-reverting behavior of real exchange rates. Even with the seemingly short-sample period of the modern float, the DF-GLS test results show that parity reversion can still be unveiled if a sufficiently powerful test is applied. A common view among previous studies of PPP is that the post-Bretton Woods period is far too short to reveal any significant parity reversion. Our results illustrate that the time span limitation may not necessarily be insurmountable if an efficient test is used.

To show the relative power of the ADF and DF-GLS tests for the sample size of our data series ($T = 261$), a simple Monte Carlo experiment is conducted using a data generating process of a stationary AR(1) model with its root equal to 0.97 (the approximate dominant root value corresponding to the relatively more persistent series in our data set). Power estimates are all obtained based on 20 000 replications in simulation. The ADF and DF-GLS tests with no time trend are considered, with their lag parameters being commonly set equal to 3. The two tests show noticeable differences in power: the power estimates of the ADF and DF-GLS test are given, respectively, by 0.20 and 0.32 for a 5% test size and 0.35 and 0.51 for a 10% test size (approximate standard errors of these estimates are calculated to be

0.002). The Monte Carlo results confirm the potential gain in power from using the DF-GLS test instead of the ADF test in the presence of persistent dynamics.

The differences between the ADF and DF-GLS results can also be understood in the context of relative efficiency. Stock (1994) analyzes the relative efficiency of various unit-root tests, including the ADF test and the DF-GLS test. In examining the behavior of two given tests of the same hypothesis against a sequence of local alternatives, the Pitman measure of relative efficiency is computed as a ratio of the sample sizes needed for these tests to produce asymptotically the same power for that sequence. Stock (1994) observes that to achieve 50% power, for example, the ADF test has an efficiency ratio of 1.91 relative to an optimal test, such as the DF-GLS test. This implies that to achieve the same level of power, the ADF test requires approx. 90% more observations than what the DF-GLS test needs. In other words, the use of the DF-GLS test has a similar effect on power as almost doubling the sample size for the ADF test. Under this perspective, the approach adopted in this study can be viewed as complementary to the sample-extension approach considered in long-sample PPP studies.

The empirical findings here are also comparable to those reported by recent panel studies of PPP (Frankel and Rose, 1996; Oh, 1996; Lothian, 1997; Papell, 1997). Although the present study, which uses univariate testing procedures, apparently differs from the panel studies in terms of the statistical approach, all of them yield a common basic result, namely, the mostly negative findings from previous studies can primarily be the result of the poor power of the usual tests. When efficient unit-root tests are used, much supportive evidence of PPP reversion can still be uncovered over the current float.

To further illustrate our analysis in comparison with those of panel studies, the panel unit-root analysis is performed directly on our recent float data, using the panel test devised by Levin and Lin (1992) and using Germany, the US and the UK as the base country, alternately. The Levin–Lin test has been commonly used and discussed in panel studies of PPP [see Levin and Lin (1992) for the exact formulation of the test]. The panel test statistics are calculated to be -3.426 for the Germany-based system, -3.680 for the US-based system and -3.299 for the UK-based system. The critical values for these test statistics are computed by the Monte Carlo method for our specific sample size; they are -3.958 for a 10% test and -4.307 for a 5% test. In contrast to our earlier results, the panel test is not able to reject the unit-root hypothesis for our recent float data, regardless of whether Germany, the US or the UK is used as the base country. These results do not contradict those reported in the other panel studies, nevertheless. Papell (1997) observes that the statistical results from panel tests are sensitive to the panel size and the country grouping. In particular, the panel test is found not to perform well in small panels, such as EMS countries. Hence, the efficient univariate testing approach applied in this study can be a useful, complementary alternative to the panel approach, especially for analyzing data for small panels. Indeed, the robustness of panel unit-root tests has been called into question. The ability to reject a unit root in individual series of real exchange rates is thus attractive. Such rejection results for univariate series can augment and reinforce those PPP findings from panel studies.

6. Concluding remarks

The return of flexible exchange rates since early 1973 has notably spawned tremendous interest in the PPP theory. The PPP relationship has been a central building block for many models of exchange rate determination (e.g. the Frenkel–Bilson and the Dornbusch–Frankel models). Also, it has often been used to provide a yardstick for evaluating the level of an exchange rate in policy discussion. The simplicity of the specification and analytical basis of the PPP relationship, as Lothian (1997) observes, may have contributed to its popular use. The PPP theory can be viewed as the open-economy extension of the quantity theory, and it posits that nominal disturbances will have no permanent effects on the real exchange rate. The persistence in PPP deviations thus reflects the nature of underlying disturbances.

Although frequent short-run departures from PPP are commonly recognized, many economists continue to hold the view that PPP, as a long-run relationship, will prevail. The faith in the PPP doctrine is, however, seriously challenged by the empirical evidence from the recent float experience. Most studies of the recent float period report evidence of a unit root in the real exchange rate, implying that shocks to real exchange rates have infinitely long-lived effects. This typical finding suggests that permanent real disturbances are the predominant source of real exchange rate fluctuations during the recent float and that theoretical or empirical modeling of the underlying determinants of PPP deviations should focus primarily on real factors. Furthermore, the real exchange rate appears to exhibit rather different behavior outside the recent float period. When other earlier historical periods are examined, much supportive evidence for long-run PPP can be found, indicating that permanent real shocks are not an important source of variability in the pre-float periods.

In finding supportive evidence of PPP reversion in the recent float data directly, the results of this study complement those from long historical data and support that, at least in terms of the existence of parity reversion, the behavior of the real exchange rate differs not much between the current float and other pre-float periods (Lothian and Taylor, 1996). In view of the unit-root rejection results, moreover, theoretical and empirical models emphasizing real factors as the principal determinants of real exchange rate dynamics will seriously overstate the importance of real disturbances relative to nominal disturbances during the current float. In a different context of economic forecasting, the findings of no unit root in the real exchange rate should also have implications for time-series forecasting since level-stationary and difference-stationary models will generate different dynamics of real exchange rates and thus different forecasts, especially long-horizon forecasts. Jorion and Sweeney (1996) and Lothian and Taylor (1996), for example, exploit the implications for forecasting in their PPP analyses.

Table 4
Finite sample critical values for the DF-WS test

Coefficients and statistics	Without trend		With trend	
	10%	5%	10%	5%
τ_0	−2.245 (0.002)**	−2.547 (0.002)**	−2.908 (0.002)**	−3.200 (0.003)**
τ_1			−5.073 (0.568)**	−5.482 (0.665)**
τ_2	−77.183 (6.331)**	−118.453 (7.103)**	−68.919 (23.360)**	−123.163 (27.558)**
δ_1	−1.339 (0.067)**	−1.618 (0.079)**	−2.614 (0.089)**	−2.720 (0.103)**
δ_2	−3.853 (0.495)**	−3.643 (0.592)**	1.439 (0.677)**	1.347 (0.797)*
R^2	0.979	0.979	0.988	0.988
$\hat{\sigma}_\epsilon$	0.014	0.017	0.016	0.018
Mean $\lvert\hat{\epsilon}\rvert$	0.011	0.013	0.013	0.014

Notes. The response surface regression is represented by Eq. (A.1). Corresponding heteroskedasticity-consistent standard errors for coefficient estimates are in parentheses. Statistical significance is indicated by a single asterisk (*) for the 10% level and a double asterisk (**) for the 5% level. $\hat{\sigma}_\epsilon$ represents the standard error of the regression. Mean $\lvert\hat{\epsilon}\rvert$ gives the mean absolute error of the response surface predictions vs. estimated critical values from simulations.

Acknowledgements

The authors are grateful to Janice Breuer, Menzie Chinn, Mark Watson and two anonymous referees for useful comments and suggestions. Research assistance from Eiji Fujii is appreciated. Any remaining errors are ours.

Appendix A: Finite sample critical values

Although the asymptotic distributions are known for all the unit-root test statistics, practical applications necessarily deal with finite samples. Finite sample critical values for the DF-GLS$^\tau$ and DF-GLS$^\mu$ tests can both be computed from response surface equations estimated by Cheung and Lai (1995b). A similar analysis is employed to obtain estimates of critical values for the DF-WS test. The analysis provides a general result for other users of the DF-WS test in different empirical applications. The experimental design covers 232 possible combinations of (T, p), with $T = \{35, 37, 40, 42, 45, 47, 50, 52, 55, 57, 60, 62, 65, 67, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, 400, 450, 500\}$ and $p = \{0, 1, 2, 3, 4, 5, 6, 7\}$. The data-generating process follows a random walk, which can produce reasonably reliable estimates of critical values for more general data processes, except for those with significant moving-average dependence. For each given (T, p) pair, finite sample critical values are computed as quantiles directly from the empirical distribution in simulations, using 30 000 replications.

After much experimentation, the following response surface equation of a second-order polynomial form is found to fit the data particularly well:

$$CR_{T,p} = \tau_0 + \sum_{j=1}^{2} \tau_i (1/T)^i + \sum_{j=1}^{2} \delta_j (p/T)^j + \epsilon_{T,p} \tag{A.1}$$

where $CR_{T,p}$ is the critical value estimate for a sample size T and lag p; and $\epsilon_{T,p}$ is the error term. The second summation term captures the incremental effects of the lag order. Since both $1/T$ and $p/T \to 0$ as $T \to \infty$, the intercept term gives an estimate of the asymptotic critical value.

Table 4 contains the results of response surface regressions. Various measures of data fit are computed, including the squared multiple correlation coefficient (R^2), standard error of regression ($\hat{\sigma}_\epsilon$), and mean absolute error (mean $|\hat{\epsilon}|$). The response surface equations appear all able to fit the data very well, in view of the high goodness of fit as measured by R^2. Both measures of $\hat{\sigma}_\epsilon$ and mean $|\hat{\epsilon}|$ are satisfactorily small in all regressions.

References

Abuaf, N., Jorion, P., 1990. Purchasing power parity in the long run. J. Finance 45, 157–174.

Ardeni, P.G., Lubian, D., 1991. Is there trend reversion in purchasing power parity? Eur. Econ. Rev. 35, 1035–1055.

Banerjee, A., Lumsdaine, R.L., Stock, J.H., 1992. Recursive and sequential tests of the unit-root and trend-break hypotheses: Theory and international evidence. J. Bus. Econ. Stat. 10, 271–287.

Baxter, M., Stockman, A.C., 1989. Business cycles and the exchange rate system. J. Monet. Econ. 23, 377–400.

Breuer, J.B., 1994. An assessment of the evidence on purchasing power parity. In: Williamson, J. (Ed.), Estimating Equilibrium Exchange Rates. Institute for International Economics, Washington, pp. 245–277.

Campbell, J.Y., Mankiw, N.G., 1987. Are output fluctuations transitory? Q. J. Econ. 102, 857–880.

Cheung, Y.W., Lai, K.S., 1993a. A fractional cointegration analysis of purchasing power parity. J. Bus. Econ. Stat. 11, 103–112.

Cheung, Y.W., Lai, K.S., 1993b. Long-run purchasing power parity during the recent float. J. Int. Econ. 34, 181–192.

Cheung, Y.W., Lai, K.S., 1995a. Lag order and critical values of the augmented Dickey–Fuller test. J. Bus. Econ. Stat. 13, 277–280.

Cheung, Y.W., Lai, K.S., 1995b. Lag order and critical values of a modified Dickey–Fuller test. Oxford Bull. Econ. Stat. 57, 411–419.

Culver, S.E., Papell, D.H., 1995. Real exchange rates under the gold standard: Can they be explained by the trend break model? J. Int. Mon. Fin. 14, 539–548.

Diebold, F.X., Husted, S., Rush, M., 1991. Real exchange rates under the gold standard. J. Polit. Econ. 99, 1252–1271.

Edison, H.J., 1987. Purchasing power parity in the long run: A test of the dollar/pound exchange rate (1890–1978). J. Mon. Cr. Banking 19, 376–387.

Edison, H.J., Gagnon, J.E., Melick, W.R., 1997. Understanding the empirical literature on purchasing power parity: The post-Bretton Woods era. J. Int. Mon. Fin. 16, 1–17.

Edison, H.J., Klovland, J.T., 1987. A quantitative reassessment of the purchasing power parity hypothesis: Evidence from Norway and the United Kingdom. J. Appl. Econometr. 2, 309–333.
Elliott, G., Rothenberg, T.J., Stock, J.H., 1996. Efficient tests for an autoregressive unit root. Econometrica 64, 813–836.
Enders, W., 1988. ARIMA and cointegration tests of PPP under fixed and flexible exchange rate regimes. Rev. Econ. Stat. 70, 504–508.
Fisher, E.O'N., Park, J.Y., 1991. Testing purchasing power parity under the null hypothesis of cointegration. Econ. J. 101, 1476–1484.
Frankel, J.A., 1986. International capital mobility and crowding-out in the US economy: imperfect integration of financial markets or goods markets? In: Hafer, R.W. (Ed.), How Open is the US Economy? Lexington Books, Lexington, pp. 33–67.
Frankel, J.A., Rose, A.K., 1996. A panel project on purchasing power parity: mean reversion within and between countries. J. Int. Econ. 40, 209–224.
Froot, K.A., Rogoff, K., 1995. Perspectives on PPP and long-run real exchange rates. In: Grossman, G., Rogoff, K. (Eds.), Handbook of International Economics, vol. 3, North-Holland, New York, pp. 1647–1688.
Glen, J.D., 1992. Real exchange rates in the short, medium, and long run. J. Int. Econ. 33, 147–166.
Johansen, S., Juselius, K., 1990. Maximum likelihood estimation and inference on cointegration — With applications to the demand for money. Oxford Bull. Econ. Stat. 52, 169–210.
Jorion, P., Sweeney, R.J., 1996. Mean reversion in real exchange rates: Evidence and implications for forecasting. J. Int. Mon. Fin. 15, 535–550.
Levin, A., Lin, C.F., 1992. Unit root tests in panel data: asymptotic and finite-sample properties. University of California at San Diego, Discussion Paper 92–23.
Lothian, J.R., 1997. Multi-country evidence on the behavior of purchasing power parity under the current float. J. Int. Mon. Fin. 16, 19–35.
Lothian, J.R., Taylor, M.P., 1996. Real exchange rate behavior: The recent float from the perspective of the past two centuries. J. Polit. Econ. 104, 488–509.
Mussa, M., 1986. Nominal exchange rate regimes and the behavior of real exchange rates: Evidence and implications. Carnegie-Rochester Conf. Ser. Pub. Pol. 25, 117–213.
Obstfeld, M., 1993. Model trending real exchange rates. University of California at Berkeley Center for International and Development Economics Research Working Paper, C93–011.
Obstfeld, M., Rogoff, K., 1995. Exchange rate dynamics redux. J. Polit. Econ. 103, 624–660.
O'Connell, P.G.J., 1996. The overvaluation of purchasing power parity. Manuscript, Harvard University.
Oh, K.Y., 1996. Purchasing power parity and unit root tests using panel data. J. Int. Mon. Fin. 15, 405–418.
Pantula, S.G., Gonzalez-Farias, G., Fuller, W.A., 1994. A comparison of unit-root test criteria. J. Bus. Econ. Stat. 12, 449–459.
Papell, D.H., 1997. Searching for stationarity: purchasing power parity under the current float. J. Int. Econ. 43, 313–332.
Park, H.J., Fuller, W.A., 1995. Alternative estimators and unit root tests for the autoregressive process. J. Time Ser. Anal. 16, 415–429.
Perron, P., Vogelsang, T.J., 1992. Nonstationarity and level shifts with an application to purchasing power parity. J. Bus. Econ. Stat. 10, 301–320.
Rogoff, K., 1992. Traded goods consumption smoothing and the random walk behavior of the real exchange rate. Bank Jpn. Monet. Econ. Stud. 10, 1–29.
Rogoff, K., 1996. The purchasing power parity puzzle. J. Econ. Lit. 34, 647–668.
Stock, J.H., 1994. Unit roots, structural breaks and trends. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, vol. 4. North-Holland, New York, pp. 2739–2284.
Taylor, M.P., McMahon, P.C., 1988. Long-run purchasing power parity in the 1920s. Eur. Econ. Rev. 32, 179–197.