Journal of International Economics 46 (1998) 281–312

Journal of
INTERNATIONAL
ECONOMICS

The behavior of real exchange rates during the post-Bretton Woods period

Mark P. Taylor[a,b,c,*], Lucio Sarno[c]

[a]*University College, Oxford University, Oxford OX1 4BH, UK*
[b]*Centre for Economic Policy Research, 90–98 Goswell Road, London EC1V 7DB, UK*
[c]*Department of Economics, Oxford University, Manor Road, Oxford OX1 3UL, UK*

Received 17 December 1996; received in revised form 14 July 1997; accepted 15 September 1997

Abstract

Since standard tests for mean reversion in real exchange rates may lack power with data spanning the recent float, researchers have employed more powerful multivariate tests. Such tests may, however, reject joint non-stationarity when just one of the processes is stationary. We suggest another test, easily constructed and with a known limiting distribution, whose null hypothesis is violated only when all of the processes in question are stationary. We investigate the finite-sample properties of both types of test by Monte Carlo simulation. Finally, we apply the tests to real exchange rates among the G5 over the recent float.

Keywords: Real exchange rates; Purchasing power parity; Multivariate unit root test; Test power; Monte Carlo simulation

JEL classification: F31

1. Introduction

The purchasing power parity (PPP) hypothesis states that national price levels expressed in a common currency should be equal. Equivalently, strict PPP implies that movements in the nominal exchange rate should be proportional to the ratio of national price levels or that the real exchange rate should be constant. PPP has

*Corresponding author. Tel.: +44-1865-281281; fax: +44-1865-281282; e-mail: mark.taylor@economics.oxford.ac.uk

PII: S0022-1996(97)00054-8

variously been viewed as a theory of exchange rate determination, as a short-run or long-run equilibrium condition, and as an efficient arbitrage condition in either goods or asset markets (Officer, 1982; Dornbusch, 1987; Taylor, 1995; Froot and Rogoff, 1995; Rogoff, 1996).

The professional literature on PPP has a long history (Officer, 1982). Prior to the recent float, the professional consensus appeared to support the existence of a varying but fairly stable real exchange rate over long periods of time (e.g. Friedman and Schwartz, 1963; Gaillot, 1970). The prevailing orthodoxy of the early 1970s, however, assumed the much stronger proposition of continuous PPP (e.g. Frenkel, 1976; Frenkel and Johnson, 1978). In the mid to late 1970s, in the light of the very high variability of real exchange rates after the major exchange rates were allowed to float, this extreme position was largely abandoned (Frenkel, 1981). Subsequently, studies published mostly in the 1980s, which could not reject the hypothesis of random walk behavior in real exchange rates (e.g. Roll, 1979; Adler and Lehmann, 1983; Piggott and Sweeney, 1985), and related work which failed to find cointegration between nominal exchange rates and relative prices (e.g. Taylor, 1988; Corbae and Ouliaris, 1988; Enders, 1988; Mark, 1990) further reduced professional confidence in PPP and led to the widespread belief that it was of little or no use empirically (e.g. Dornbusch, 1988).[1]

A possible rationalisation of the widespread failure to reject non-stationarity of real exchange rates, suggested by a number of authors, is that the span of available data for the recent floating rate period alone may simply be too short to provide any reasonable degree of test power in the normal statistical tests for non-stationarity (Frankel, 1989; Lothian and Taylor, 1997; Hakkio and Rush, 1991). Accordingly, researchers have sought to remedy this by increasing the sample period under investigation, (e.g. Frankel, 1986, 1989; Edison, 1987; Abuaf and Jorion, 1990; Kim, 1990; Lothian, 1990; Hakkio and Joines, 1990; Diebold et al., 1991[2]; Lothian and Taylor, 1996). As noted by Frankel and Rose (1996) and others, however, the long samples required to generate a reasonable level of statistical power with standard stationarity tests may be unavailable for many currencies and may potentially be inappropriate because of regime changes. While some authors, notably Lothian and Taylor (1996), have argued that reliable inferences can be drawn by extending the data across exchange rate regimes – at least concerning the stability of the first moments of real exchange rate series – others remain skeptical of this view. A number of authors, including Baxter and

[1] At the same time, however, some authors have reported results supporting long-run purchasing power parity for certain historical periods such as the interwar period (Taylor and McMahon, 1988), the International Gold Standard (Diebold et al., 1991) or the 1950s Canadian float (Choudhry et al., 1991) or under special circumstances such as high inflation episodes (McNown and Wallace, 1989) or among the exchange rates of member countries of the European Monetary System (Cheung and Lai, 1995a), thereby creating a puzzle as to why PPP failed to hold for the major exchange rates during the recent float.

[2] Diebold et al. (1991) apply fractional integration techniques – see also Cheung and Lai (1993a).

Stockman (1989); Mussa (1986); Frankel (1989); Hegwood and Papell (1998), argue that the statistical properties of the real exchange rate appear to vary strongly across nominal exchange rate regimes.[3] To settle the issue of whether the real exchange rate has behaved in a mean-reverting fashion over the post-Bretton Woods period would therefore seem to require inference based on data for the recent float alone.

A second approach has therefore been taken by some researchers, involving the use of panel data on exchange rates over relatively shorter periods of time. Flood and Taylor (1996), for example, analyze a panel of annual data on 21 industrialized countries over the floating rate period and find strong support for mean reversion towards long-run purchasing power parity by regressing 5, 10 and 20 year average exchange rate movements on average inflation differentials with the US. Frankel and Rose (1996) analyze a very large panel of annual data on 150 countries in the post World War II period and also find evidence of mean reversion similar to that evident in studies of long time series. In an influential paper Abuaf and Jorion (1990) develop a multivariate unit root test based on systems estimation of autoregressive processes for a set of real exchange rate series, and use this to reject the joint null hypothesis of non-stationarity of a number of real exchange rates for the recent floating rate period. Panel data methods have also been applied to this issue by, inter alios, Wei and Parsley (1995); Wu (1996); Oh (1996); O'Connell (1998); Papell (1998).

In the present paper, we seek to contribute to this literature in a number of ways. Firstly, we provide some further evidence on panel unit root tests of this kind, by calculating the finite sample empirical distribution of a multivariate augmented Dickey-Fuller (MADF) statistic while allowing for higher-order serial correlation in real exchange rates and relaxing the assumption that the sum of the autoregressive coefficients are identical across the panel under the alternative hypothesis.

Secondly, however, we point out and illustrate through Monte Carlo simulations an important potential pitfall in the interpretation of multivariate unit root tests of this kind. The pitfall is simply this: the null hypothesis in panel unit root tests is usually that all of the series under consideration are realizations of unit root processes. Thus, the null hypothesis will be violated even if only one of the real exchange rate series in the panel is in fact stationary. Hence, although such multivariate tests may be informative under certain conditions, they may also be relatively uninformative since rejection of the null hypothesis will in general not help the researcher in determining how many of the series under consideration are stationary. We show, inter alia, that multivariate unit root tests of this kind may lead to a very high probability of rejection of the joint null hypothesis of non-stationarity when there is a single stationary process among a system of

[3] Relatedly, Grilli and Kaminsky (1991) argue that real exchange rate behavior varies more with the historical period per se than across exchange rate regimes.

otherwise unit root processes, even when the root of the single stationary process is close to the unit circle.

Thirdly, therefore, we investigate by Monte Carlo methods the finite-sample empirical performance of a multivariate test in which the null hypothesis is that at least one of the series in the panel is a realization of a unit root process. This null hypothesis is only violated if all of the series are in fact realizations of stationary processes.[4] Moreover, the test procedure we suggest is now widely available to researchers since it simply involves a special application of Johansen's (1988) maximum likelihood procedure for testing for the number of cointegrating vectors in a system. A further attractive property of this test which we demonstrate is that, in the special case we examine – ie. under the null hypothesis that at least one of the series is a realization of a unit root process – it has a known limiting $\chi^2(1)$ distribution. We compute finite-sample critical values for this test but we also show that the finite sample empirical distribution is quite close to the asymptotic distribution in sample sizes exceeding about 100, corresponding approximately to the number of quarterly observations currently available for the recent float.[5]

The remainder of the paper is set out as follows. In Section 2 we briefly outline the PPP hypothesis and the long-run properties of real exchange rates which it implies. In Section 3 we outline two multivariate unit root tests based on a generalization of the augmented Dickey-Fuller test statistic and of the Johansen maximum likelihood cointegration procedure respectively. In Section 4 we discuss some preliminary data analysis and univariate unit root tests on four dollar real exchange rates over the floating rate period. In Section 5 we report Monte Carlo evidence on the two multivariate tests described in Section 3. In Section 4 we employ these tests on quarterly dollar real exchange rates constructed using nominal exchange rates and either relative consumer price indices or relative GDP deflators among the G5 countries over the recent floating rate period. In Section 7

[4]Given, that is, the maintained hypothesis – common to all test procedures of this kind – that the series are realizations of either $I(1)$ or $I(0)$ processes (where an $I(d)$ process can be thought of as one which must be differenced d times before it becomes stationary). Evidence of explosive behavior in real exchange rates – for example positive Dickey-Fuller statistics – can easily be checked for prior to applying multivariate test procedures.

[5]One can distinguish between panel data studies in which the number of time series is relatively large (e.g. Frankel and Rose, 1996) and those where the number of series is relatively small (e.g. Abuaf and Jorion, 1990; Jorion and Sweeney, 1996). Formally, if N is the number of time series in the panel and T is the sample size, different econometric results follow according to whether T is assumed fixed and N is assumed to be relatively large, or whether N is assumed fixed and T is assumed to be relatively large, or whether both N and T are assumed to be large (which would normally require an additional assumption such as that N/T is small) – see Im et al. (1997). The conceptual issues raised in this paper relate to tests of long-run PPP based on both small and large panels, ie. tests involving any of the standard assumptions regarding N and T. The particular tests we investigate, however, are applicable primarily to relatively small systems of real exchange rates, such as is the case commonly encountered in testing for long-run PPP for a group of industrialized countries' exchange rates over the recent floating rate period.

we report further empirical and Monte Carlo work based on real exchange rates among the same countries constructed using producer price indices; this section is designed as a check on the generality of the simulation results derived earlier as well as an investigation of real exchange rates constructed using price indices containing a smaller proportion of non-tradables. A final section summarizes and concludes.

2. Long-run purchasing power parity

PPP may be examined through the real exchange rate since the logarithm of the real exchange rate, q_t, can be defined as the deviation from PPP:

$$q_t \equiv s_t + p_t^* - p_t \tag{1}$$

where s_t denotes the logarithm of the nominal exchange rate (domestic price of foreign currency) observed at time t and p_t and p_t^* are the logarithms of the domestic and foreign price levels respectively.

While q_t may be subject to considerable short-run variation, a necessary condition for PPP to hold in the long run is that the real exchange rate q_t be stationary over time, not driven by permanent shocks. If this is not the case, then the nominal exchange rate and the price differential will permanently tend to deviate from one another. This is the rationale for applying non-stationarity tests to real exchange rate data as a means of testing for long-run purchasing power parity.

It should be noted, moreover, that there are good economic reasons why real exchange rate movements should contain permanent components, particularly where the price indices used in the construction of the real rate contain both tradables and non-tradables. The well known Harrod-Balassa-Samuelson effect (Harrod, 1933; Balassa, 1964; Samuelson, 1964), for example, implies that relatively fast growing countries may have a tendency to have higher real exchange rates based on relative consumer price indices (see e.g. Froot and Rogoff, 1995; Rogoff, 1996; Obstfeld and Rogoff, 1996). As noted by Rogoff (1996), however, while there is reasonably strong evidence suporting the Harrod-Balassa-Samuelson effect between very rich and very poor countries, its empirical relevance for the long-horizon time-series behaviour of real exchange rates among industrialised countries remains a matter of debate, possibly because of the effects of technology diffusion.[6]

[6]Equilibrium models of the exchange rate (Stockman, 1980; Lucas, 1982), in which the real exchange rate is driven primarily by persistent real shocks such as shifts in tastes and technology, have also been used to rationalize persistence in real exchange rate movements. On the other hand, one might expect that relative real shocks affecting the real exchange rate between industrialised countries may have a higher mean-reverting component because of technology diffusion and other catch-up effects. Moreover, empirical tests of the implications of such models – for example that the real exchange rate should be invariant to nominal exchange rate regimes – has not by and large been favourable (Taylor, 1995).

Besides the issue of traded-goods productivity bias, other arguments may be adduced at a theoretical level to suggest why real exchange rates may heave persistent components. For example, differences in aggregate growth rates across countries may induce permanent changes in real exchange rates through preferences if consumers' Engle curves bend towards non-traded goods. Obstfeld and Rogoff (1995) show that in the presence of sticky goods prices, monetary shocks may have a long-run effect on the real exchange rate because of the residual effects of temporary current account imbalances occasioned by short-run movements in the real exchange rate.

Overall, it is probably true to say that few economists would rule out the possibility of real long-run effects on real exchange rates altogether. In testing for long-run PPP we are, therefore, implicitly testing whether permanent real effects account for only a relatively small part of long-run real exchange rate movements. To that extent, long-run PPP becomes an empirical matter.

3. Multivariate unit root tests

In this section we outline two multivariate unit root tests. The first may be considered as an extension of previous work due to Abuaf and Jorion (1990). The second test we propose is an application of Johansen's maximum likelihood procedure for testing for the number of cointegrating vectors (Johansen, 1988, 1991) in the unusual case where the number of cointegrating vectors tested for is exactly equal to the number of time series in the system.

3.1. A multivariate augmented Dickey-Fuller test

Following the work of Fuller (1976) and Dickey and Fuller (1979, 1981), it is possible to test for a unit root in the stochastic process generating a time series q_t by estimating the auxiliary regression:

$$q_t = \mu + \sum_{j=1}^{k} \rho_j q_{t-j} + u_t$$

where the number of lags k is chosen such that the residual u_t is approximately white noise. For stationarity we require $\Sigma_{j=1}^{k} \rho_j < 1$, while if q_t is a realization of a unit root process, one should expect to find $\Sigma_{j=1}^{k} \rho_j = 1$. The augmented Dickey-Fuller test statistic is the standard '*t*-ratio' test statistic for $H_0: \Sigma_{j=1}^{k} \rho_j = 1$ and the rejection region consists of large negative values. As is well known, this statistic does not, however, follow the standard Student's t-distribution under the null hypothesis because of the theoretically infinite variance of q_t and finite-sample critical values have been computed using Monte Carlo methods by Fuller (1976) and MacKinnon (1991). This statistic is normally termed the Dickey-Fuller (DF) statistic for $k=0$ and the augmented Dickey-Fuller statistic (ADF) for $k>0$.

The first multivariate test for unit roots we propose is a multivariate analogue of the standard, single-equation augmented Dickey-Fuller test – the multivariate ADF or MADF test. Consider an $(N\times1)$ dimensional stochastic vector process generated in discrete time according to:

$$q_{it} = \mu_i + \sum_{j=1}^{k} \rho_{ij} q_{it-j} + u_{it} \tag{3}$$

for $i=1,...,N$ and $t=1,...,T$, where N denotes the number of series in the panel and T is the number of observations. The disturbances $\boldsymbol{u_t}=(u_{1t}...u_{Nt})'$ are assumed to be independently normally distributed with a possibly non-scalar covariance matrix:

$$\boldsymbol{u_t} \sim IN(0, \Lambda) \tag{4}$$

The standard, single-equation ADF unit root test would involve estimating each of the N equations separately and carrying out N individual tests of the null hypothesis:

$$\mathrm{H}_{0i}: \sum_{j=1}^{k} \rho_{ij} - 1 = 0 \tag{5}$$

For situations where the root of each of the individual autoregressive process is close to but less than unity, it is well known that univariate ADF tests may lack power.

The approach taken in this paper is to estimate equation (3) as a system of N equations, taking account of contemporaneous correlations among the disturbances,[7] and to test equation (5) jointly on all N equations:

$$\mathrm{H}_0: \sum_{j=1}^{k} \rho_{ij} - 1 = 0, \ \forall i = 1,..., N \tag{6}$$

taking the resulting Wald statistic as the MADF statistic.

The obvious way to estimate equation (3) jointly is to employ Zellner's (1962) 'seemingly unrelated' (SUR) estimator, which is basically multivariate generalized least squares (GLS) using an estimate of the contemporaneous covariance matrix of the disturbances obtained from individual ordinary least squares estimation. We can write equation (3) in matrix notation as:

$$\boldsymbol{Q} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{u} \tag{7}$$

[7] O'Connell (1998) demonstrates the importance of accounting for cross-sectional dependence among real exchange rates when testing for long-run PPP. He shows that failure to allow for contemporaneous correlation of the residuals may generate very large size distortion in panel unit root tests.

where the $NT \times 1$ vector $\boldsymbol{Q}$ is given by $\boldsymbol{Q} = (q_1' \ q_2' ... q_N)'$, q_i is a $T \times 1$ vector of observations on the i-th real exchange rate, with t-th element q_{it}; $\mathbf{Z}$ is an $NT \times N(k+1)$ block diagonal matrix with the i-th block a $T \times (k+1)$ matrix with ones in the first column and T observations on k lags of q_{it} in the remainder of the matrix; $\boldsymbol{\beta}$ is an $N(k+1) \times 1$ vector of stacked parameters for each equation; $\boldsymbol{u}$ is an $NT \times 1$ vector containing the stacked disturbances, so that[8]

$$\boldsymbol{u} \sim N(O, \boldsymbol{\Lambda} \otimes \mathbf{I}_T) \tag{8}$$

The restrictions in the null hypothesis equation (6) may then be written as:

$$\boldsymbol{\Psi\beta} - \boldsymbol{\iota} = \boldsymbol{O} \tag{9}$$

where $\boldsymbol{\Psi}$ is an $N \times N(k+1)$ block-diagonal matrix with the i-th block a $1 \times (k+1)$ row vector with zero as the first element and unity elsewhere, $\boldsymbol{\iota}$ is an $N \times 1$ vector of ones and $\boldsymbol{O}$ is an $N \times 1$ vector of zeroes. The MADF test statistic for the unit root hypothesis equation (6) is the standard Wald test statistic which may be written:

$$MADF = \frac{(\boldsymbol{\iota} - \boldsymbol{\Psi}\hat{\boldsymbol{\beta}})'\{\boldsymbol{\Psi}[\mathbf{Z}'(\hat{\boldsymbol{\Lambda}}^{-1} \otimes \mathbf{I}_T)\mathbf{Z}]^{-1}\boldsymbol{\Psi}'\}(\boldsymbol{\iota} - \boldsymbol{\Psi}\hat{\boldsymbol{\beta}})N(T-k-1)}{(\boldsymbol{Q} - \mathrm{Z}\hat{\boldsymbol{\beta}})'(\hat{\boldsymbol{\Lambda}}^{-1} \otimes \mathbf{I}_T)(\boldsymbol{Q} - \mathrm{Z}\hat{\boldsymbol{\beta}})} \tag{10}$$

where $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Lambda}}$ are consistent estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\Lambda}$[9]. In general, the Wald statistic for testing N restrictions has a limiting χ^2 distribution with N degrees of freedom under the null hypothesis being tested. In the present case, however, its distribution is unknown because of the theoretically infinite variance of the processes generating the real exchange rate series under the null hypothesis equation (6). Its finite-sample empirical distribution can, however, be calculated by Monte Carlo simulation.

Abuaf and Jorion (1990) suggest a similar multivariate test, based on estimation of a first-order autoregressive equation for each individual real exchange rate, with the first-order autocorrelation coefficient constrained to be equal across exchange rates. Their proposed test statistic is then the ratio of the estimated common parameter minus one to its estimated standard error. The multivariate ADF statistic we propose can thus be viewed as a generalisation of the Abuaf-Jorion approach, to allow for higher order serial correlation in real exchange rates and to allow the

---

[8]Here and throughout the paper, $\mathbf{I}_j$ denotes the $j \times j$ identity matrix, for $j = T$ or N.

[9]In fact, a consistent estimate of β as well as Λ could be obtained from OLS applied individually to each equation since, unlike the case in Abuaf and Jorion (1990), we actually impose no cross-equation restrictions. Given a non-diagonal contemporaneous residual covariance matrix, however, the SUR estimator will be a more efficient estimator of β than OLS and so the finite-sample performance of the MADF should be better using SUR rather than individual OLS estimates.

sum of the autoregressive coefficients to vary across exchange rates under the alternative hypothesis.[10]

3.2. The Johansen likelihood ratio test

Johansen (1988, 1991) suggests a maximum likelihood procedure for testing for the number of cointegrating vectors in a multivariate context. Engle and Granger (1987) demonstrate that, among a system of N I(1) series, there can be at most $N-1$ cointegrating vectors. Thus, if we reject the hypothesis that there are less than N cointegrating vectors among N series, this is equivalent to rejecting the hypothesis of non-stationarity of all of the series. Equivalently, the only way there can be N distinct cointegrating vectors among N series is if each of the series is I(0) and so is itself a cointegrating relationship.[11]

The Johansen likelihood ratio (JLR) test for cointegration is based on the rank of a long-run multiplier matrix in a vector autoregressive system. Consider the data generating process of an $N\times 1$ vector process $\boldsymbol{Q}_t$ – which may be assumed to generate realizations of N real exchange rates at time t – in vector autoregressive (VAR) form:

$$\boldsymbol{Q}_t = \boldsymbol{\Pi}_1 \boldsymbol{Q}_{t-1} + \cdots + \boldsymbol{\Pi}_k \boldsymbol{Q}_{t-k} + \boldsymbol{\mu} + \boldsymbol{\omega}_t, \quad t = 1, 2, \ldots, T \tag{11}$$

where the $\boldsymbol{\Pi}_i$s are $(N\times N)$ matrices of parameters, $\boldsymbol{\mu}$ is an $N\times 1$ vector of constants and ω_t is an $N\times 1$ vector of white noise errors. This VAR system can be reparameterized into the error correction form:

$$\Delta\boldsymbol{Q}_t = \boldsymbol{\Gamma}_1 \Delta\boldsymbol{Q}_{t-1} + \cdots + \boldsymbol{\Gamma}_{k-1} \Delta\boldsymbol{Q}_{t-k+1} + \boldsymbol{\Gamma}_k \boldsymbol{Q}_{t-k} + \boldsymbol{\mu} + \boldsymbol{\omega}_t \tag{12}$$

where:

$$\boldsymbol{\Gamma}_i = -\mathbf{I} + \boldsymbol{\Pi}_1 + \cdots + \boldsymbol{\Pi}_i, \quad i = 1, \ldots, k \tag{13}$$

and $\boldsymbol{\Gamma}_k$ represents the long-run solution of the VAR. Indeed, $\boldsymbol{\Gamma}_k$ is an $N\times N$ matrix whose rank defines the number of distinct cointegrating vectors. To see this, note that $\boldsymbol{\Gamma}_k$ may be written:

$$\boldsymbol{\Gamma}_k = -\boldsymbol{\alpha}\boldsymbol{\gamma}' \tag{14}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are each $N\times[\mathrm{rank}(\boldsymbol{\Gamma}_k)]$ matrices. $\boldsymbol{\gamma}$ can be interpreted as the matrix

[10] Recent papers by Papell (1998) and O'Connell (1998) also allow for higher order serial correlation in tests of this kind, although these authors retain the restriction that the autoregressive coefficients are identical across the panel under both the null and alternative hypotheses. The Monte Carlo work of Papell (1998) shows that the presence of serial correlation may affect the size of panel unit root tests.

[11] Again, as in Johansen (1988, 1991); Johansen and Juselius (1990), this implicitly assumes a maintained hypothesis that the series are either I(0) or I(1).

of cointegrating parameters and $\boldsymbol{\alpha}$ as the matrix of error correction coefficients (Johansen, 1988). If, for example, each of the series is individually I(1) and no cointegrating vectors exist, then since all of the other terms in equation (12) are I(0), $\boldsymbol{\Gamma}_k$ must be the null matrix so that rank($\boldsymbol{\Gamma}_k$) = 0. To take the opposite extreme, if $\boldsymbol{\Gamma}_k$ is of full rank then the space spanned by $\boldsymbol{\gamma}$ is N-dimensional Euclidean space and contains $\mathbf{I}_N$. Hence, N cointegrating vectors can be formed, each consisting of just one of the series, which implies that each of the series must be I(0) and the vector process X_t is stationary (see e.g. Johansen and Juselius, 1990).

The rank of a matrix is equal to the number of non-zero latent roots. In the present context, stationarity of all of the processes in the vector autoregressive system is tantamount to $\boldsymbol{\Gamma}_k$ having full rank and so N non-zero latent roots. Thus, the null hypothesis of one or more non-stationary processes making up an $N \times 1$ vector process can be expressed:

$$H_0\text{: rank}(\boldsymbol{\Gamma}_k) < N \tag{15}$$

and tested against the alternative hypothesis that each of the series is stationary, or equivalently:

$$H_1\text{: } rank(\boldsymbol{\Gamma}_k) = N \tag{16}$$

Since full rank of $\boldsymbol{\Gamma}_k$ would imply that all of the latent roots are non-zero, a test of equation (15) can be based only on the smallest latent root, since rejection of the hypothesis that the smallest latent root is zero is sufficient to reject the hypothesis that $\boldsymbol{\Gamma}_k$ has less than full rank.

Following Johansen (1988, 1991), an appropriate test statistic for the null hypothesis that the smallest latent root of $\boldsymbol{\Gamma}_k$ is zero, equivalent to a likelihood ratio test of equation (15) against equation (16) can be constructed as follows. First, correct for the effects of $\mathfrak{F}_t = \{\boldsymbol{\iota}, \Delta\boldsymbol{Q}_{t-1}, \Delta\boldsymbol{Q}_{t-2}, \ldots, \Delta\boldsymbol{Q}_{t-k+1}\}$ on $\Delta\boldsymbol{Q}_t$ and $\boldsymbol{Q}_{t-k}$ (where $\boldsymbol{\iota}$ is the unit vector) by projecting each of them onto $\mathfrak{F}_t$ and retrieving the residuals, denoted $\mathbf{R}_{ot}$ and $\mathbf{R}_{kt}$ respectively. Then form the matrices $\mathbf{S}_{ij} = \mathbf{T}^{-1}\Sigma_{t=1}^{T}\mathbf{R}_{it}\mathbf{R}'_{jt}$ $(i, j = 0, k)$, and extract the smallest root, λ_N say, of the characteristic equation $|\lambda\mathbf{S}_{kk} - \mathbf{S}_{ko}\mathbf{S}_{oo}^{-1}\mathbf{S}_{ok}| = 0$. Johansen's likelihood ratio statistic is then:[12]

$$JLR = -T\ln(1 - \lambda_N) \tag{17}$$

[12]The Johansen procedure can be related to the estimated $\boldsymbol{\Gamma}_k$ in the following way. Testing for the zero roots of $|\lambda\mathbf{S}_{kk} - \mathbf{S}_{ko}\mathbf{S}_{oo}^{-1}\mathbf{S}_{ok}| = 0$ is equivalent to testing for the zero roots of $|\lambda - \mathbf{S}_{ko}\mathbf{S}_{oo}^{-1}\mathbf{S}_{ok}\mathbf{S}_{kk}^{-1}| = 0$. But, by the Frisch and Waugh (1933) theorem, $\mathbf{S}_{ok}\mathbf{S}_{kk}^{-1}$ is the multivariate least squares estimator of $\boldsymbol{\Gamma}_k$ in equation (12). By the same token, $\mathbf{S}_{ko}\mathbf{S}_{oo}^{-1}$ is the estimated matrix of coefficients of ΔQ_t resulting from applying least squares to equation (12) with $\Delta\boldsymbol{Q}_t$ and $\boldsymbol{Q}_{t-k}$ interchanged.

JLR as defined in equation (17) is a likelihood ratio statistic for one restriction ($\lambda_N = 0$). Johansen and Juselius (1990) show that JLR converges weakly to a function of Brownian motion:

$$JLR \Rightarrow \frac{\left[\int_0^1 (t - 1/2)\mathrm{d}B\right]^2}{\int_0^1 (t - 1/2)^2 \mathrm{d}t} \tag{18}$$

where B is a standard Brownian motion. Although, in general, the likelihood ratio statistics derived by Johansen (1988, 1991); Johansen and Juselius (1990) have non-standard limiting distributions, the right hand side of equation (18) is in fact distributed as $\chi^2(1)$, so that, in this special case, the Johansen likelihood ratio statistic has a standard χ^2 distribution with one degree of freedom in large samples (Johansen and Juselius, 1990).[13]

3.3. The MADF and the JLR tests compared

There is a subtle but very important difference between the null hypothesis tested by the MADF statistic equation (10) and that tested by the JLR statistic equation (17). The null hypothesis equation (6) will be violated if one or more of the series in question is a realization of an I(0) process. The null hypothesis equation (15) will, however, only be violated if all of the N series are realizations of I(0) processes. This implies that multivariate Dickey-Fuller tests should be interpreted with caution in the context of testing for non-stationarity of real exchange rates. Abuaf and Jorion (1990), for example, apply a restricted form of the MADF test to a system of ten real exchange rates and reject the null hypothesis of joint non-stationarity at the 5% level: while this may imply ten stationary real exchange rates, it may equally imply only one or two. This suggests that the JLR test statistic may provide a useful alternative or complement to the MADF or similar panel unit root tests.

Below, we investigate the power of the MADF and JLR statistics under a

[13]This result arises for two reasons. First, we are testing for the significance of only one latent root. Second, because we placed no restrictions on the constant intercept terms in the VAR equation (12), we implicitly allowed for the possibility of linear trends (Johansen and Juselius, 1990, p. 171). This is therefore a special case of the result due to West (1988), that if a linear trend is present under the null hypothesis of non-stationarity, then the usual asymptotics hold for the likelihood ratio test.

variety of assumptions concerning the number of non-stationary series in the system under consideration.

4. Preliminary data analysis and single-equation unit root tests

Quarterly data on bilateral real dollar exchange rates among the G5 countries (i.e. sterling–dollar, mark–dollar, franc–dollar and yen–dollar) for the period 1973i–1996ii were constructed from series obtained from the International Monetary Fund's International Financial Statistics (IFS) data bank.[14] Initially, we examined two real exchange rate series. These correspond to the nominal exchange rate (currency per dollar) deflated by, respectively, relative consumer price indices (CPI) and relative GDP deflators. In Section 7 below we analyze real exchange rates for these countries constructed using producer price indices. All of the price indices are based on 1990. Each of the real exchange rate series was put into natural logarithms before the econometric analysis.

The first task was to estimate univariate autoregressive equations for each of the series and to construct single-equation unit root tests. In every case, a first-order autoregressive process appeared unsatisfactory in that significant serial correlation remained in the residuals. A priori, one might expect a fourth-order autoregression to be more suitable for quarterly data, and, in fact, this turned out to be the case in terms of eliminating serial correlation of the residuals. A fourth-order autoregressive model was also preferred on the basis of the Akaike Information Criterion (Akaike, 1973), the Schwartz Information Criterion (Schwartz, 1978) or the method proposed by Campbell and Perron (1991).[15,16]

Table 1 (Panel A) lists the estimated coefficients of the fourth-order autoregressions as well as test statistics for serial correlation in the residuals and ADF test statistics for each of the real exchange rates. In all cases, in keeping with the

[14] In investigating the low-frequency characteristics of time series processes, Shiller and Perron (1985) note that the span of the data set – in terms of years – is far more important than the number of observations per se. An intuitive discussion of this point is given in Davidson and MacKinnon (1993), Chapter 20. Our choice of quarterly data should therefore make our analysis of wider interest to other researchers (since quarterly data are available for a wider range of countries than are monthly data) without a major loss of power of the tests.

[15] The Campbell-Perron method involves starting with a high-order autoregression and sequentially excluding the highest-order lag until the coefficient on the highest-order lag is statistically significant.

[16] In their Monte Carlo simulations, Cheung and Lai (1993b) show that for autoregressive processes with no moving average dependencies the Akaike Information Criterion and the Schwartz Information Criterion indicate the right lag order of a vector autoregression used for testing for cointegration in 99.86% and 99.96% of cases respectively.

literature on univariate tests for mean reversion in the major real exchange rates over the period since 1973, we were unable to reject at the 5% level the null hypothesis of non-stationarity on the basis of the single-equation ADF test statistics.

A number of authors, notably Kremers et al. (1992); Campos et al. (1996); Banerjee et al. (1993, chapter 7), have pointed out that tests for unit roots and non-cointegration based on ADF tests may implicitly impose certain common factor restrictions across the parameters of the data generating process of the time series under investigation. Accordingly, we estimated trivariate vector autoregressive systems in the (logarithms of the) nominal exchange rate and domestic and US prices for each country and for all definitions of the price level, and tested the cross-equation restrictions necessary for the system to be reduced to a univariate autoregression in the real exchange rate, using likelihood ratio tests. In no case could these restrictions be rejected at a nominal significance level of 5%.[17]

In Table 1 (Panel B) we report the standard deviations of the residuals from estimating single-equation AR(4) models for the real exchange rates, together with the contemporaneous cross-exchange rate residual correlation matrix. The likelihood ratio statistics for the null hypothesis that the residual covariance matrices are diagonal [LR(diag)] massively reject the null hypothesis, suggesting that systems estimation should yield substantial efficiency gains and that panel unit root tests applied to this data without allowing for this cross-sectional dependence would very likely be subject to substantial size distortion (O'Connell, 1998).

5. Monte Carlo simulations

The Monte Carlo experiments were based on a data generation process consisting of one to four autoregressive models, each of order four. From the descriptive statistics reported in Table 1 (Panel B) we derive the average contemporaneous covariance matrix for the AR(4) residuals, averaged across the CPI and GDP deflator adjusted real exchange rate residual covariance matrices (Table 1 Panel C). We employ this average covariance matrix in executing the Monte Carlo simulations. For each autoregressive model we took the average parameters for the estimates of the AR(4) process using the two real exchange rate series for each of the countries, as reported in Table 1, adjusting the slope

[17]Standard likelihood ratio tests were applied to the fourth-order trivariate systems. In fact, the marginal significance levels in virtually every case were very much larger then 5%. Full details are available from the authors on request.

Table 1
Preliminary data analysis

Panel A:[a] Single-equation AR(4) estimates and ADF tests on bilateral dollar exchange rate

	Lag 1	Lag 2	Lag 3	Lag 4	Intercept	Adj. R^2	$Q(27)$	DW	Sum of AR Coeffs.	ADF
UK										
CPI	1.104	−0.207	0.096	−0.097	−0.014	0.86	31.21 (0.26)	2.016	0.896	−2.407
GDPD	1.110	−0.169	0.048	−0.083	−0.018	0.88	36.12 (0.11)	2.012	0.906	−2.392
Mean	1.107	−0.188	0.072	−0.090	−0.016				0.901	
Germany										
CPI	1.247	−0.391	0.307	−0.232	−0.003	0.92	16.58 (0.94)	2.051	0.931	−2.198
GDPD	1.235	−0.327	0.241	−0.216	−0.007	0.93	17.15 (0.93)	2.052	0.933	−2.163
Mean	1.241	−0.359	0.274	−0.224	−0.005				0.932	
France										
CPI	1.328	−0.504	0.234	−0.129	−0.012	0.91	12.06 (0.99)	2.016	0.929	−2.096
GDPD	1.308	−0.456	0.186	−0.111	−0.008	0.91	11.84 (0.99)	2.012	0.927	−2.128
Mean	1.318	−0.480	0.210	−0.120	−0.010				0.928	
Japan										
CPI	1.330	−0.493	0.287	−0.163	−0.018	0.95	27.04 (0.46)	2.011	0.961	−1.678
GDPD	1.304	−0.431	0.205	−0.125	−0.016	0.94	22.96	2.011	0.953	−1.703
Mean	1.317	−0.462	0.246	−0.144	−0.017				0.957	

Panel B:[b] Residual standard deviations and contemporaneous correlations:

	UK	*Germany*	*France*	*Japan*
CPI				
Standard deviations	0.052	0.046	0.044	0.049
Correlations:				
UK	1.000			
Germany	0.631	1.000		
France	0.639	0.907	1.000	
Japan	0.457	0.565	0.554	1.000
LR(diag) = 249.88				
GDPD				
Standard deviation	0.054	0.049	0.047	0.051
Correlations:				
UK	1.000			
Germany	0.633	1.000		
France	0.644	0.900	1.000	
Japan	0.460	0.563	0.553	1.000
LR(diag) = 245.47				

Panel C:[c] Average covariance matrix

	UK	*Germany*	*France*	*Japan*
UK	0.283 E−2			
Germany	0.159 E−2	0.224 E−2		
France	0.111 E−2	0.120 E−2	0.208 E−2	
Japan	0.172 E−2	0.215 E−2	0.127 E−2	0.254 E−2

[a]Notes: CPI and GDPD denote the real exchange rate deflated by relative consumer price indices and implicit GDP deflators respectively. Adj. R^2 denotes the degrees-of-freedom adjusted coefficient of determination; $Q(27)$ is the Ljung-Box test statistic for residual serial correlation up to 27 lags (P-values in parentheses); DW denotes the Durbin-Watson test for first-order serial correlation; and ADF is the augmented Dickey-Fuller statistic (i.e. the 't-ratio' for the slope coefficients to sum to unity).

[b]Notes: Covariance and correlation matrices were constructed using the residuals from the AR(4) regressions reported in Panel A LR(diag) is a likelihood ratio statistic for the null hypothesis that the off-diagonal elements of the relevant covariance matrix are zero, and has a limiting χ^2 distribution with six degrees of freedom if the true covariance matrix is diagonal.

[c]Notes: Entries correspond to the arithmetic average of the corresponding entries in the two covariance matrices for the residuals from the AR(4) regressions given in Panel A.

parameters in order to adjust their sum.[18] Specifically, the systems were based on the following data generating process:

$$q_{1t} = -0.016 + (1.107 + \delta_1)q_{1t-1} + (-0.188 + \delta_1)q_{1t-2} + 0.072q_{1t-3} - 0.090q_{1t-4} + u_{1t} \tag{19}$$

$$q_{2t} = -0.005 + (1.241 + \delta_2)q_{2t-1} + (-0.359 + \delta_2)q_{2t-2} + 0.274q_{2t-3} - 0.224q_{2t-4} + u_{2t} \tag{20}$$

$$q_{3t} = -0.010 + (1.318 + \delta_3)q_{3t-1} + (-0.480 + \delta_3)q_{3t-2} + 0.210q_{3t-3} - 0.120q_{3t-4} + u_{3t} \tag{21}$$

$$q_{4t} = -0.017 + (1.317 + \delta_4)q_{4t-1} + (-0.462 + \delta_4)q_{4t-2} + 0.246q_{4t-3} - 0.144q_{4t-4} + u_{4t} \tag{22}$$

where $(u_{1t}\ u_{2t}\ u_{3t}\ u_{4t})' \sim N(0, \Lambda)$ with Λ as given in Table 1 (Panel C), and the δ_i denote the relevant adjustments to the parameters. In order to generate the critical values for the test statistics, for example, we needed to generate the replications with the coefficients summing to unity in each case. Thus, we set $\delta_1 = (1-0.901)/2$, $\delta_2 = (1-0.932)/2$, $\delta_3 = (1-0.928)/2$ and $\delta_4 = (1-0.957)/2$. To take another example, in order to examine the behavior of the statistics under the alternative hypothesis where all of the autoregressive processes had coefficients summing to 0.95, we set set $\delta_1 = (0.95-0.901)/2$, $\delta_2 = (0.95-0.932)/2$, $\delta_3 = (0.95-0.928)/2$ and $\delta_4 = (0.95-0.957)/2$.

All of the Monte Carlo results discussed in this paper were constructed using 5000 replications in each experiment, with identical random numbers across experiments (Hendry, 1984). The simulations were executed for a number of different sample sizes ($T = 25$, 50, 75, 100, 200, 300, 500). At each replication we started with the first four initial values of each of the artificial series set to zero. We then generated a sample size of $105 + T$ ($T = 25$, 50, 75, 100, 200, 300, 500) and

[18] It might be argued that a separate data generating process should be formulated for each data set (ie. CPI and GDP deflator adjusted real exchange rates), rather than averaging across the covariance matrix or across the estimated autoregressive parameters. While the approach adopted in this paper has the advantage of making the results of the Monte Carlo simulations more readily comprehensible by reducing the quantity of results to assimilate, it is clearly important to check for the generality of the results. In order to check the robustness of our Monte Carlo results to slight changes in the assumed data generating process, we therefore performed a number of safeguards. First, we carried out Monte Carlo simulations with data generating processes corresponding more specifically to each of the two data sets, for experiments corresponding to approximately a third of the data used to construct Figs. 1–3. In every case, the results were qualitatively unaffected, affecting at most the second decimal place of the rejection frequencies. In addition, we investigated how the empirical critical values for the MADF are affected by quite wide adjustments to the assumed covariance matrix (see footnote 26). In Section 7, we also report the results of a case study of real exchange rates constructed using producer price indices, the Monte Carlo simulations for which are based on a data generating process calibrated on those particular series.

discarded the first 105 observations, leaving a sample of size T for the analysis.[19] For the simulated I(0) processes – ie. those with autoregressive coefficients summing to less than unity – this should reduce the dependence of the results on the initialization. By definition, however, the simulated I(1) processes are long-memory in nature and so are unavoidably contingent upon the initialization. An initialization of zero for the log real exchange rate does, however, seem reasonable.

5.1. The MADF test: Monte Carlo evidence

Table 2 reports the 5% empirical critical values for the MADF test with $N=4$ and the various sample sizes considered.[20]

In Fig. 1 we have plotted the percentage of rejections of the null hypothesis of joint non-stationarity when each of the series is a realization of a stationary process, for various true values of the sum of the autoregressive coefficients.[21] As this reveals, the MADF test is quite powerful even when the sum of the autoregressive parameters is very close to unity. With a sample size of 100, for example, the null hypothesis of non-stationarity is rejected in around 30% of the replications when each of the series is generated by a process with a root of 0.99. For roots of 0.975, the rejection rate rises to around 60%, to close to 90% for roots

Table 2
MADF: empirical critical values at the 5% level

$T=25$	$T=50$	$T=75$	$T=100$	$T=200$	$T=300$	$T=500$
30.7115	21.2993	18.5062	16.8701	14.5119	13.5531	12.2813

Note: The critical values correspond to a fourth-order system. T is the sample size.

[19] A sample size of 25 was not used for analysis of the JLR statistic because of the very low degrees of freedom in a fourth-order VAR with four lags with only 25 observations. Sample sizes of 300 and 500 were only used for analysis of the MADF test because of the high computational expense of Monte Carlo analysis with samples of this size with the Johansen procedure.

[20] In order to see how sensitive the results were to the assumed covariance matrix of the innovations, we also generated the critical values assuming two alternative covariance matrices: a half correlations matrix, in which the covariances between residual series were halved while keeping variances unchanged; and a diagonal covariance matrix, in which cross-correlations are all set to zero. The critical values obtained from experiments using the alternative covariance matrices (available from the authors on request), however, differed only slightly from the critical values generated from the simulations employing the full historical matrix.

[21] Full tabulations of all of the empirical power functions plotted in Figs. 1–3 are available from the authors on request.

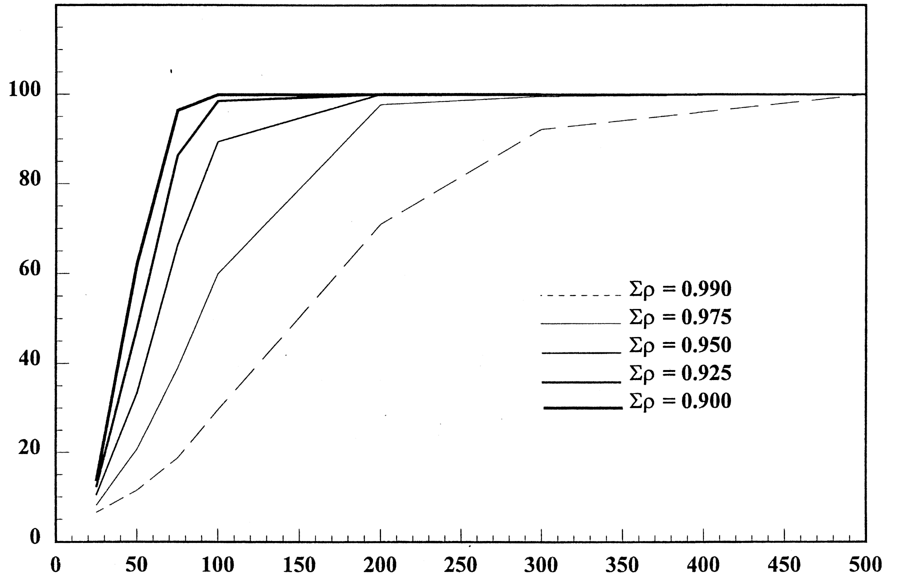


Fig. 1. MADF: estimated power function.

of 0.95, nearly 99% for roots of 0.925 and close to 100% for roots of 0.9 or less.[22]

We also employed response surface analysis to obtain approximate finite-sample critical values for the MADF test statistic. Response surfaces reduce the specificity of the Monte Carlo analysis by permitting interpolation for points in the experimental design space that were not simulated.[23] After considerable experimentation, the following form was found to work extremely well:

$$C(Tj) = \sigma_0 + \sigma_1 T(j)^{-1} + \sigma_2 T(j)^{-2} + \sigma_3 T(j)^{-3} + \text{errors} \tag{23}$$

where $C(Tj)$ is the 5% critical value for a sample size T obtained from the jth experiment. In practice, the response surface we propose is similar to the one used by MacKinnon (1991) with the addition of one higher-order polynomial term, which was found to improve significantly the explanatory power. Using ordinary least squares estimation yielded:

[22]For purposes of comparison, we also calculated the empirical power function for a univariate ADF test applied to the first equation in the system (assuming an AR(4) process), using 5% critical values calculated from the response surface results given in MacKinnon (1991), although we do not report this in order to conserve space. This exercise demonstrated, however, that the MADF test is very much more powerful than the univariate ADF test. For a sample size of 100 and a single root of 0.990, for example, the ADF statistic has a rejection frequency of just a little over 6%, compared to nearly 30% for the MADF statistic when there are four processes with the same root (Fig. 1). For roots of 0.95 the MADF statistic has a rejection frequency of nearly 90%, compared with just over 17% for the univariate ADF.

[23]Useful references on response surface methodology are Hendry (1984) and Ericsson (1991).

$$C(Tj) = \underset{(0.29)}{11.12} + \underset{(85.57)}{720.82}\, T(j)^{-1} - \underset{(5859.88)}{15646.71}\, T(j)^{-2} + \underset{(98476.59)}{246820.8}\, T(j)^{-3} + \text{errors}$$

$$R^2 = 0.99\ \text{DW} = 2.12\ \text{LM}(1) = 0.15\ \text{RESET}(1) = 0.12$$
$$\text{JB}(2) = 0.34\ \text{BP}(1) = 0.24\ \text{ARCH}(1) = 0.41 \tag{24}$$

where figures in parenthesis below estimated coefficients are estimated standard errors, R^2 is the coefficient of determination, DW is the Durbin-Watson statistic, LM(1) is a Lagrange multiplier test statistic for first-order serial correlation, RESET is Ramsey's (1969) test statistic for functional form misspecification, JB is the Jarque and Bera (1980) test for normality of the residuals, BP is the Breusch and Pagan (1979) test for heteroskedasticity, ARCH is a test for first-order autoregressive conditional heteroskedasticity. Higher-order powers of the regressor were found to be insignificant. These results suggest that MacKinnon's response surface with one higher-order term provides a reasonable approximation to the finite-sample critical values in this particular case.[24][25]

In Fig. 2 we plot the MADF rejection frequencies of the null hypothesis of joint non-stationarity of the processes when, respectively, one, two or three of them have a root less than unity. As Fig. 2 makes clear, the MADF test remains quite powerful under these circumstances. For example, when two of the processes are I(1) and two are stationary (Fig. 2B), the rejection frequency is around 23% at the 5% significance level and a sample size of 100 when the roots of the stationary processes are each equal to 0.99, rising quickly to over 45% for roots of 0.975 and to over 80% for roots of 0.95.[26][27]

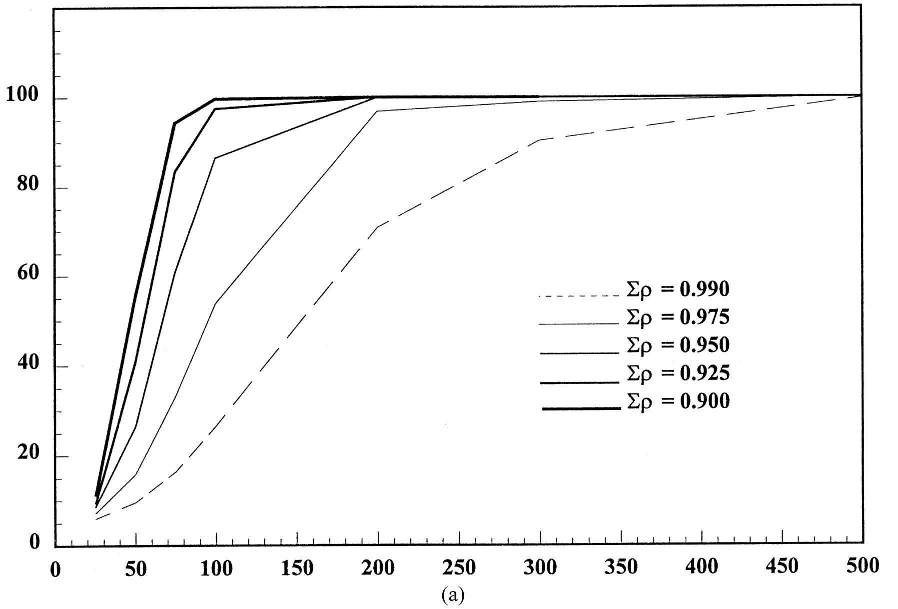
---

[24] Cheung and Lai (1995b) also show that the lag order, in addition to the sample size, may affect the finite-sample behavior of the ADF and hence point the importance of correcting for the effect of lag order in applying the ADF statistic. While we keep the lag order fixed in the present analysis, the lag order may affect the behavior of the MADF in a similar way as the ADF and deserves future investigation.

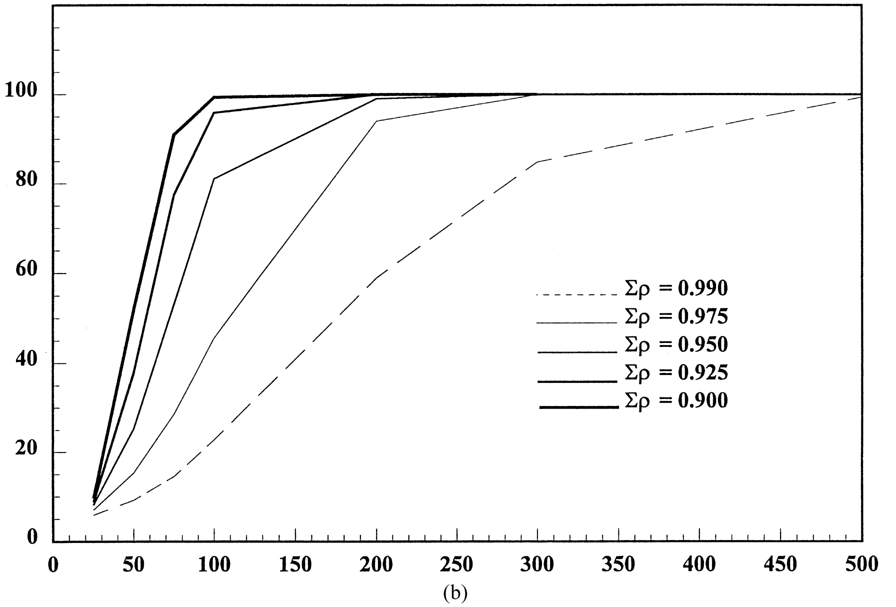
[25] We also considered the Reinsel and Ahn (1988) approximation for the response surface, but – consistently with Cheung and Lai (1995b) – this was not found to yield an improved fit relative to the higher-order MacKinnon response surface regression reported above.

[26] Following Abuaf and Jorion (1990), we also investigated the possibility that the small-sample behavior of the MADF test may be affected, in terms of both critical values and power, by the presence of autoregressive conditionally heteroskedastic (ARCH) effects in the disturbances. Using ARCH(1) and generalized ARCH, GARCH(1,1) parameters for the data generation processes based on actual sample estimates, we found that the power of the MADF tests was little affected. As the parameters were arbitrarily increased towards one, however, both the power and the actual test size increased slightly. These results, indicating that the test does not appear too sensitive to the presence of conditional heteroskedasticity, are in line with those reported by Abuaf and Jorion (1990).

[27] We also performed Monte Carlo experiments for the MADF with smaller systems – ie. with $N=2$ and $N=3$. For $N=3$ we omitted Eq. (22) from the data generating process, and for $N=2$ we omitted Eq. (21) and Eq. (22) (as well as the corresponding rows and columns of Λ in each case). The major noteworthy characteristic of these experiments was a noticeable drop in power as the number of series in the system is reduced.

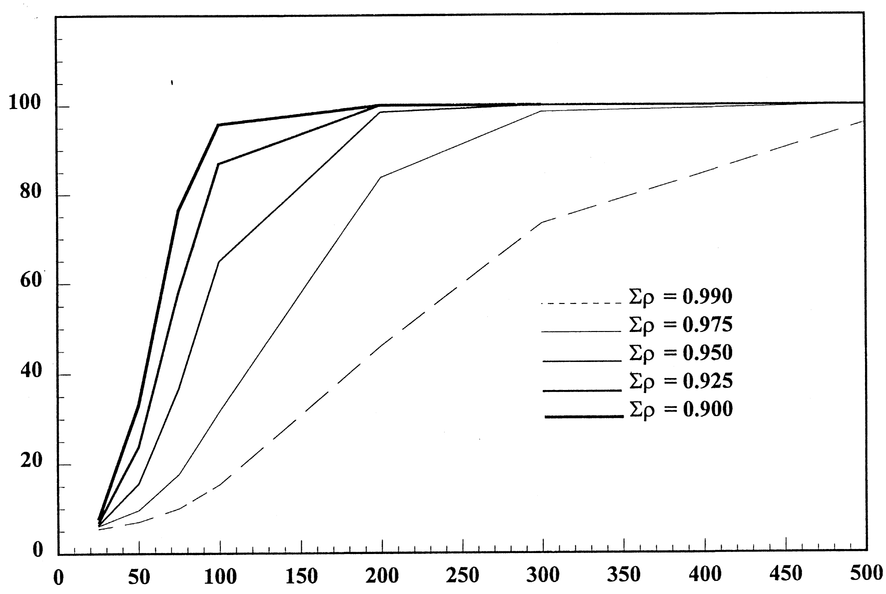


(a)



(b)

Fig. 2. (a) MADF: estimated power function with three stationary processes and one unit root processes; (b) MADF: estimated power function with two stationary processes and two unit root processes; (c) MADF: estimated power function with one stationary process and three unit root processes.



(c)

Fig. 2. (*continued*)

At the same time, however, Fig. 2 illustrates the potential pitfall in the use – or rather the interpretation – of the MADF or similar panel unit root tests in testing for long-run PPP: rejection of the null hypothesis does not necessarily indicate that all of the processes in the system under consideration are stationary. For $N=4$, $T=100$, with three unit root processes in the system together and with just one stationary process with a root of 0.95, for example, the null hypothesis is rejected at the 5% level in 65% of the replications (Fig. 2C). For one stationary process with a root of 0.9 together with three unit root processes and a sample size of 100, the rejection frequency is in excess of 95%. For larger sample sizes, the rejection frequencies when there is only a single stationary process are even greater.

5.2. The JLR test: Monte Carlo evidence

As we showed above, the JLR statistic does in fact have a limiting $\chi^2(1)$ distribution under the null hypothesis of less than full rank of the long-run matrix. Nevertheless, we generated its small-sample empirical distribution, since Cheung and Lai (1993b) show that there may be substantial finite sample bias toward rejection of the null hypothesis in Johansen likelihood ratio statistics.

In Table 3, therefore, we report the critical values, obtained from executing 5000 simulations, for the JLR statistic under the null hypothesis that the rank of the long-run matrix is less than full. As noted above, less than full rank of this

Table 3
JLR: average empirical critical values

$T=50$	$T=75$	$T=100$	$T=200$
5.5065	4.3133	4.0686	3.9712

Notes: Using a fourth-order system, for each sample size, four experiments were performed in which, respectively all four, the first three, the first two, and the first of the four autoregressive equations in the data generating process had coefficients summing to unity while the remainder were set equal to the estimated values given in Table 1 (Panel A). Each experiment involved 5000 replications and the critical value was taken as the 95th percentile. The average empirical critical value was then taken and is given in the table. T is the sample size.

matrix corresponds to the case of one or more non-stationary processes in the system. Accordingly, the critical values were calculated under all possible cases which satisfy the null hypothesis – ie. for all values of the number of non-stationary processes from four down to one – and the arithmetic average taken as the appropriate entry of Table 3. For the smaller samples, the estimated finite sample critical values are quite large, compared to the corresponding critical values from the $\chi^2(1)$ distribution. For the larger sample sizes of 100 or more, however, the critical values are quite close to those of the $\chi^2(1)$ distribution. For a sample size of $T=100$, for example, the 5% critical value for a system with $N=4$ is 4.0686. This compares with the 5% critical value form the $\chi^2(1)$ distribution of 3.84 and the large-sample critical value for JLR estimated by Johansen and Juselius (1990) (Table A1) of 3.962. This suggests that in sample sizes of 100 or more – corresponding roughly to the number of quarterly observations available for the floating rate period – researchers could assume that JLR follows a $\chi^2(1)$ distribution under the null hypothesis with only slight size distortion.

Reinsel and Ahn (1988) suggest a finite-sample scaling factor adjustment of $T/(T-Nk)$ to the asymptotic critical values of Johansen test statistics in order to obtain their finite-sample counterparts. Although the Monte Carlo study of Cheung and Lai (1993b) suggests that the Reinsel-Ahn adjustment does not yield unbiased estimates of the finite sample critical values, we thought it worth examining this hypothesis in the present situation where the limiting distribution is a known $\chi^2(1)$ distribution. Accordingly, following Cheung and Lai (1993b), we fitted by ordinary least squares a response surface of the form

$$\frac{\zeta(Tj)}{\zeta[\chi^2(1)]} = \alpha_0 + \alpha_1 \frac{T}{(T-Nk)} + \text{errors} \tag{25}$$

where $\zeta(Tj)$ is the finite-sample 5% level critical value for the j-th experiment and $\zeta[\chi^2(1)]$ is the corresponding critical value from the $\chi^2(1)$ distribution. This yielded:

$$\frac{\zeta(Tj)}{\zeta[\chi^2(1)]} = -0.149 + 1.096\frac{T}{(T-Nk)}$$
$$\quad [0.335] \quad [0.282]$$

$$R^2 = 0.95 \ \mathrm{DW} = 2.21 \ \mathrm{LM}(1) = 0.29$$
$$[0.59]$$

$$\mathrm{RESET}(1) = 0.17 \ \mathrm{JB}(2) = 0.84 \ \mathrm{BP}(1) = 0.14$$
$$[0.67] \qquad [0.66] \qquad [0.70]$$

$$\mathrm{ARCH}(1) = 0.11 \ \chi^2(\alpha_0 = 0, \alpha_1 = 1)1.23$$
$$[0.74] \qquad [0.54] \tag{26}$$

using the same notation as in Section 5.1, and where the final test statistic relates to a linear Wald test that the intercept is zero and the slope coefficient is unity. Figures in parenthesis below estimated coefficients are estimated standard errors and in brackets below test statistics are marginal significance levels. Higher-order powers of the regressor were found to be insignificant. These results suggest that, in this particular case, the Reinsel-Ahn adjustment may provide a reasonable approximation to the finite-sample critical values.

In Fig. 3 we plot the empirical power function of the JLR test when all of the processes are integrated of the same order, using the finite-sample 5% critical

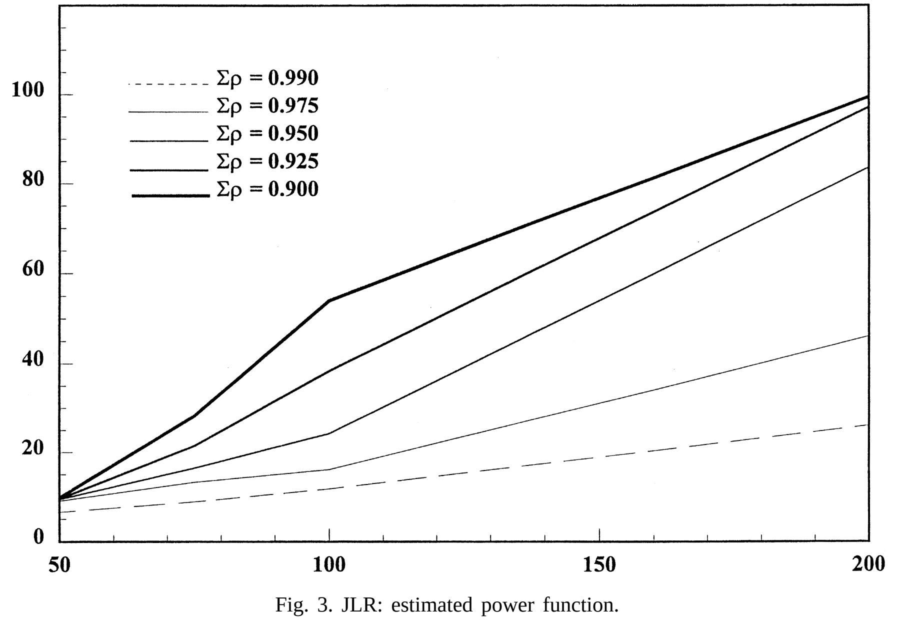


Fig. 3. JLR: estimated power function.

value. Note that these rejection frequencies are not directly comparable to those discussed above for the MADF test because the null hypotheses for the two tests are quite different. Nevertheless, the JLR test does appear to be moderately powerful. For $N=4$, $T=100$ and at a significance level of 5%, for example, the rejection frequency is around 12% for roots of 0.99, rising to just over 16% for roots of 0.975, to just under 25% for roots of 0.95, to about 38% for roots of 0.925 and to 54% for roots of 0.9.[28]

6. Empirical results for CPI-adjusted and GDP deflator-adjusted real exchange rates

In Table 4 we report the results of applying the MADF and JLR tests to the four real dollar exchange rates (dollar–sterling, dollar–mark, dollar–franc and dollar–yen) using quarterly data for the period 1973i through 1996ii, constructed using relative consumer price indices and relative GDP deflators. We used lag lengths of four for each of the autoregressions (in the construction of the MADF test statistic) and in the vector autoregression (in the construction of the JLR test statistic).

The MADF test (Panel A) rejects the null hypothesis of joint non-stationarity at the 5% significance level for both types of real exchange rates (i.e. deflated either by relative CPIs or relative GDP deflators), thereby implying that at least one of the series in each of the systems is a realization of a stationary process.

Applying the JLR test (Panel B), we easily reject the null hypothesis that the long-run impact matrix has less than full rank when we consider the real exchange rates deflated by relative CPIs, implying that all the series in question are realizations of stationary processes. When we consider the real exchange rates

Table 4
Empirical results

Panel A: MADF test statistics		
Countries	*CPI*	*GDPD*
UK, GE, FR, JA	26.5497	26.5774
Panel B: JLR test statistics		
Countries	*CPI*	*GDPD*
UK, GE, FR, JA	5.8851	3.7712

Notes: In Panel A, the null hypothesis is that all four real exchange rate series are realizations of unit root processes, the alternative hypothesis is that at least one of them is a realization from a stationary process. The 5% critical value, taken from Table 2, is 16.8701.
In Panel B the null hypothesis is that at least one of the four real exchange rate series is a realization of a unit root process, the alternative hypothesis is that all of them are realizations of stationary processes. The 5% critical value, taken from Table 3, is 4.0686.

[28]This compares to rejection frequencies calculated for a univariate ADF statistic (not reported) of around 6%, 10%, 17%, 28% and 42% respectively.

constructed using relative GDP deflators, however, we are not able to reject the null hypothesis at the 5% level. In both cases, the JLR test results are qualitatively unaffected whether we use the finite-sample critical values given in Table 3, the relevant percentiles of the $\chi^2(1)$ distribution, or the asymptotic critical values calculated by Johansen and Juselius (1990), (Table A1).

Taken together, therefore, the MADF and JLR test results imply the following. For the CPI-adjusted real exchange rates, we reject the hypothesis that each of the four series is generated by an I(1) process (Panel A). For the same real exchange rates, we can also reject the null hypothesis that at least one of them is generated by a non-stationary process (Panel B). Hence, the strong implication is that they are each realizations of stationary processes over the floating exchange rate period.

For the GDP deflator-adjusted real exchange rates, however, while we can reject at the 5% level the hypothesis that each of the four series is generated by an I(1) process (Panel A), we are unable to reject, at the same significance level, the null hypothesis that at least one of them is generated by a non-stationary process (Panel B). The JLR test thus indicates the need for caution in interpreting the MADF test result applied to the GDP-adjusted real exchange rates: the most we can say is that at least one of them appears to be generated by a stationary process over the floating rate period.

The difference in the test results applied to the two sets of real exchange rates is perhaps not surprising since real exchange rates constructed using relative CPIs may be viewed as more appropriate for testing PPP than those constructed using relative GDP deflators, since GDP deflators will typically be constructed using a much larger proportion of non-tradable goods prices (Froot and Rogoff, 1995; Rogoff, 1996; Taylor and Sarno, 1998, Chapter 3).

7. A Case Study: producer price indices[29]

As a check on the robustness of the simulation results discussed above to the specific data generating processes assumed, we also investigated the mean-reverting behavior of real dollar exchange rates among the G5 constructed using producer price indices (PPIs). Since PPIs cover a higher proportion of tradables goods prices than either CPIs or GDP deflators, one might also expect long-run PPP to hold more strongly using these indices to construct the real exchange rate series.[30] Quarterly data on PPIs were obtained from the IFS data bank for the sample period 1980i–1996ii (66 data points), since the PPI series for France was only available from 1980i onwards.

As in the case of the the CPI-adjusted and the GDP deflator-adjusted real

[29]This section was included at the suggestion of an anonymous referee.

[30]Keynes (1932) and McKinnon (1971) argue that the Harrod-Balassa-Samuelson effect is likely to be more pronounced in real exchange rates constructed using CPIs rather than PPIs.

exchange rate series, a first-order autoregressive model appeared unsatisfactory for each of the real exchange rate series in that significant serial correlation remained in the residuals. Elimination of the residual serial correlation led to the choice of a fourth-order model in every case, and this lag length was also optimal according to the other selection criteria considered, namely the Akaike Information Criterion, the Schwartz Information Criterion and the Campbell-Perron method.

Table 5 (Panel A) lists the estimated coefficients as well as ADF test statistics and residual diagnostics. In all cases, we were unable to reject, at the 5% level, the null hypothesis of non-stationarity on the basis of the single-equation ADF test statistics.

The contemporaneous covariance matrix for the AR(4) residuals, given in Table 5 (Panel B), demonstrates strong cross-sectional effects in the data, which is confirmed formally on the basis of a likelihood ratio statistic for the diagonality of the covariance matrix. Using this estimated covariance matrix and the estimated parameters adjusted to fit the null hypothesis of four unit roots, as described in Section 5, we then constructed the MADF and JLR 5% critical values for a sample size of 66 and 5000 replications.[31]

The 5% empirical critical value for the MADF test is 18.7894, which is in the range between the critical values given in Table 2 for samples of $T=50$ (21.2993) and $T=75$ (18.5062) as generated from the average covariance matrix of CPI-adjusted and GDP deflator-adjusted real exchange rate residuals. In fact, using the data generation process described in Section 5, we estimated the 5% critical value for a sample size of $T=66$ at 18.9834. Using our estimated response surface for the MADF critical values, Eq. (24), the estimated critical value would be 19.3080.

The 5% critical value for the JLR statistic was estimated at 4.6856, which again is in the range between the critical values, given in Table 3, for samples of $T=50$ (5.5065) and $T=75$ (4.3133) as generated from the average covariance matrix given in Table 1 (Panel C). Using the data generating process described in Section 5, we estimated the JLR 5% critical value for a sample size of $T=66$ at 4.6834.

The MADF test statistic calculated on the actual PPI-adjusted real exchange rates for the period 1980i–1996ii is 19.7432 which, compared to the critical value of 18.7894, enables us to reject the null hypothesis of joint non-stationarity at the 5% level, implying that at least one of the PPI-adjusted real exchange rate series in the system is a realization of a stationary process.

The JLR test computed on the same system of PPI-adjusted real exchange rate series is, however, 4.0969 which, compared to the 5% critical value of 4.6856, does not enable us to reject the null hypothesis that the long-run impact matrix has less than full rank. In fact, the empirical marginal significance level of a value of the test statistic of 4.0969 is 6.64%,[32] so that while we were unable to reject the

[31] Following our previous practice, we initalized the first four observations to zero, generated $105+66$ observations and discarded the first 105.

[32] i.e. the percentage of experiments which generated a larger value of the test statistic under the null hypothesis.

Table 5
Estimates using PPI-adjusted real exchange rates

Panel A[a]: Single-equation AR(4) estimates and ADF tests on bilateral dollar exchange rate

	Lag 1	Lag 2	Lag 3	Lag 4	Intercept	Adj. R^2	Q(27)	DW	Sum of AR Coeffs.	ADF
UK	1.148	−0.282	0.132	−0.070	0.005	0.88	21.95 (0.58)	1.992	0.928	−1.619
Germany	1.243	−0.279	0.095	−0.114	0.005	0.93	13.04 (0.96)	2.056	0.945	−1.568
France	1.291	−0.421	0.175	−0.115	0.014	0.91	12.57 (0.97)	2.040	0.930	−1.763
Japan	1.263	−0.422	0.297	−0.181	−0.008	0.94	7.99 (0.99)	1.986	0.957	−1.334

Panel B: Covariance matrix

	UK	*Germany*	*France*	*Japan*
UK	0.264 E−2			
Germany	0.178 E−2	0.235 E−2		
France	0.167 E−2	0.216 E−2	0.226 E−2	
Japan	0.116 E−2	0.151 E−2	0.139 E−2	0.199 E−2
LR(diag) = 328.48				

[a]Notes: Adj. R^2 denotes the degrees-of-freedom adjusted coefficient of determination; $Q(27)$ is the Ljung-Box test statistic for residual serial correlation up to 27 lags (P-values in parentheses); DW denotes the Durbin-Watson test for first-order serial correlation; and ADF is the augmented Dickey-Fuller statistic (i.e. the 't-ratio' for the slope coefficients to sum to unity).

[b]Notes: Covariance and correlation matrices were constructed using the residuals from the AR(4) regressions reported in Panel A. LR(diag) is a likelihood ratio statistic for the null hypothesis that the off-diagonal elements of the covariance matrix are zero, and has a limiting χ^2 distribution with six degrees of freedom if the true covariance matrix is diagonal.

null hypothesis at the 5% level, it could be rejected at the 7% level. Given that we were able to reject the null hypothesis using the JLR test applied to the real exchange rate series constructed using consumer price indices, which would cover a higher proportion of non-tradables than the producer price indices, this suggests that the marginal inability to reject the null hypotheis at the 5% level using the PPI-adjusted real exchange rates may be due to a loss of power because of the smaller sample size.

8. Conclusion

In this paper we have provided a number of insights into multivariate tests of long-run purchasing power parity. In particular, while we have shown that panel unit root tests may be quite powerful, they must be interpreted with caution since rejection of the null hypothesis of joint non-stationarity of a group of real exchange rates may be due to as few as one of the exchange rate series under investigation being generated by a stationary process. For a sample size of around 100, for example, we found that the presence of a single stationary process in a system together with three unit root processes led to rejection, at the 5% level, of the joint null hypothesis of non-stationarity in about 65% of simulations when the root of the stationary process was as large as 0.95, and on more than 95% of occasions when the single stable root was 0.9 or less.

We therefore suggested an alternative or complementary multivariate test of non-stationarity where the null hypothesis is not that all of the series are generated by non-stationary processes, but, rather, that at least one of the series is generated by a non-stationary process. This null hypothesis will only be violated if all of the series in question are realizations of stationary processes.

Applying both of these test procedures to bilateral real dollar exchange rate series for the G5 countries during the post Bretton Woods period, we found strong evidence of mean reversion in real exchange rates constructed using consumer price indices.

For exchange rates constructed GDP deflators, while the panel unit root test indicated rejection of the joint hypothesis of non-stationarity, our suggested complementary procedure indicated that this may be due to only a sub-sample of the series being stationary. Similar remarks apply to the results obtained using real exchange rates constructed using producer price indices, although here a further complication arose because of the loss of test power concomitant with the necessity of using a smaller sample size.

However, the evidence that CPI-adjusted real exchange rates among the G5 are apparently mean reverting over the floating rate period is, by itself, an important finding of our research, corroborating other recently emerging evidence that long-run PPP may hold after all. Indeed, it seems that the profession's confidence in long-run PPP, having been low for a number of years, may itself be mean reverting.

Acknowledgements

This paper was begun while Mark Taylor was on the staff of the International Monetary Fund. We are grateful to three anonymous referees and the managing editor, Andrew Rose, for helpful and constructive comments on an earlier version of this paper. The authors have also benefited from helpful comments from and discussions with Robert Flood, Alan Taylor, Hashem Pesaran, Richard Clarida, Ron Smith, Paul O'Connell, Nelson Mark, Martin Evans, Axel Weber and David Papell, as well as other conference participants at the Centre for Economic Policy Research Conference on Market Integration and Real Exchange Rates, held under the auspices of the Program for International Economic Studies, Georgetown University, 2–3 May 1997. Any errors which remain are solely the responsibility of the authors.

References

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

Adler, M., Lehmann, B., 1983. Deviations from purchasing power parity in the long run. Journal of Finance 38, 1471–1487.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Second International Symposium on Information Theory. Akademini Kiado, Budapest, pp. 267–281.

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

Banerjee, A., Dolado, J.J., Galbraith J.W., Hendry, D.F., 1993. Co-integration, error correction, and the econometric analysis of non-stationary data. Oxford University Press, Oxford.

Baxter, M., Stockman, A.C., 1989. Business cycles and the exchange rate system. Journal of Monetary Economics 23, 377–400.

Breusch, T.S., Pagan, A.R., 1979. A simple test for heteroskedasticity and random coefficient variation. Econometrica 47, 1287–1294.

Campbell, J.Y., Perron, P., 1991. Pitfalls and opportunities: what macroeconomists should know about unit roots. In: Blanchard, O.J., Fisher, S. (Eds.), NBER Macroeconomic Annuals 6. MIT Press, Cambridge, MA.

Campos, J., Ericsson, N.R., Hendry, D.F., 1996. Cointegration tests in the presence of structural breaks. Journal of Econometrics 70, 187–200.

Cheung, Y.W., Lai, K.S., 1993. A fractional cointegration analysis of purchasing power parity. Journal of Business Economics and Statistics 11, 103–112.

Cheung, Y.W., Lai, K.S., 1993. Finite-sample sizes of Johansen's likelihood ratio tests for cointegration. Oxford Bulletin of Economics and Statistics 55, 313–328.

Cheung, Y.W., Lai, K.S., 1995. Purchasing power parity under the european monetary system. Journal of International Money and Finance 14, 179–189.

Cheung, Y.W., Lai, K.S., 1995. Lag order and critical values of the augmented Dickey-Fuller test. Journal of Business and Economic Statistics 13, 277–280.

Choudhry, T., McNown, R., Wallace, M., 1991. Purchasing power parity and the canadian float in the 1950s. Review of Economics and Statistics 73, 558–563.

Corbae, D., Ouliaris, S., 1988. Cointegration and tests of purchasing power parity. Review of Economics and Statistics 70, 508–511.

Davidson, R., MacKinnon, J.G., 1993. Estimation and inference in econometrics. Oxford University Press, New York and Oxford.
Dickey, D., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74, 427–431.
Dickey, D., Fuller, W.A., 1981. Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica 49, 1057–1072.
Diebold, F., Husted, S., Rush, M., 1991. Real exchange rates under the gold standard. Journal of Political Economy 99, 1252–1271.
Dornbusch, R., 1987. Purchasing power parity. In: Eatwell, J., Milgate, M., Newman, P. (Eds.), The New Palgrave: A Dictionary of Economics. Macmillan, London; Stockton, New York, pp. 1075–1085.
Dornbusch, R., 1988. Real exchange rates and macroeconomics: a selective survey. National Bureau of Economic Research Working Paper No. 2775.
Edison, H.J., 1987. Purchasing power parity in the long run. Journal of Money, Credit and Banking 19, 376–387.
Enders, W., 1988. ARIMA and cointegration tests of PPP under fixed and flexible exchange rate regimes. Review of Economics and Statistics 70, 504–508.
Engle, R.F., Granger, C.W.J., 1987. Cointegration and error correction: representation. Estimation and Testing, Econometrica 55, 251–276.
Ericsson, N.R., 1991. Monte Carlo methodology and the finite sample properties of instrumental variables statistics for testing nested and non-nested hypotheses. Econometrica 59, 1249–1277.
Flood, R.P., Taylor, M.P., 1996. Exchange rate economics: what's wrong with the conventional macro approach? In: Frankel, J.A., Galli, G., Giovannini, A. (Eds.), The Microstructure of Foreign Exchange Markets. Chicago University Press, Chicago, for NBER.
Frankel, J.A., 1986. International capital mobility and crowding out in the US economy: imperfect integration of financial economics or of goods markets? In: Hafer, R. (Ed.), How Open is the US Economy? Lexington Books, Lexington, MA.
Frankel, J.A., 1989. Zen and the art of modern macroeconomics: the search for perfect nothingness. In: Haraf, W., Willett, T. (Eds.), Monetary Policy for a Volatile Global Economy. American Enterprise Institute, Washington, DC.
Frankel, J.A., Rose, A.K., 1996. A panel project on purchasing power parity: mean reversion within and between countries. Journal of International Economics 40, 209–224.
Frenkel, J.A., 1976. A monetary approach to the exchange rate: doctrinal aspects and empirical evidence. Scandinavian Journal of Economics 78, 200–224.
Frenkel, J.A., 1981. The collapse of purchasing power parities during the 1970s. European Economic Review 16, 145–165.
Frenkel, J.A., Johnson, H.G., 1978. The Economics of Exchange Rates. Addison Wesley, Reading MA.
Friedman, M., Schwartz, A.J., 1963. A Monetary History of the United States: 1867–1960. Princeton University Press, Princeton NJ, for the National Bureau of Economic Research.
Frisch, R., Waugh, F.V., 1933. Partial time regressions as compared with individual trends. Econometrica 1, 387–401.
Froot, K.A., Rogoff, K., 1995. Perspectives on PPP and long run real exchange rates. In: Rogoff, K., Grossman, G. (Eds.), Handbook of International Economics. North Holland, Amsterdam.
Fuller, W.A., 1976. Introduction to Statistical Time Series. Wiley, New York.
Gaillot, H.J., 1970. Purchasing power parity as an explanation of long term changes in exchange rates. Journal of Money Credit and Banking 2, 348–357.
Grilli, V., Kaminsky, G., 1991. Nominal exchange rate regimes and the real exchange rate: evidence for the United States and Great Britain, 1885-1986. Journal of Monetary Economics 27, 191–212.
Hakkio, C.S., Rush, M., 1991. Cointegration: how short is the long run? Journal of International Money and Finance 10, 571–581.
Hakkio, C.S., Joines, D.H., 1990. Real and Nominal Exchange Rates Since 1919. Federal Reserve Bank of Kansas City Working Paper.

Harrod, R., 1933. International Economics. James Nisbet and Cambridge University Press, London.

Hegwood, N.D., Papell, D.H., 1998. Quasi Purchasing Power Parity. International Journal of Finance and Economics 3, 279–289.

Hendry, D.F., 1984. Monte Carlo experimentation in econometrics. In: Griliches, Z., Intrilligator, M.D. (Eds.), Handbook of Econometrics. North Holland, Amsterdam, New York and Oxford, pp. 939–976.

Im, K.S., Pesaran, M.H., Shin, Y., 1997. Testing for Unit Roots in Heterogeneous Panels. Mimeo, Department of Applied Economics, University of Cambridge (forthcoming Econometrica).

Jarque, C.M., Bera, A.K., 1980. Efficient tests for normality, heteroskedasticity and serial independence of regression residuals. Economics Letters 6, 255–259.

Johansen, S., 1988. Statistical analysis of cointegrating vectors. Journal of Economic Dynamics and Control 12, 231–254.

Johansen, S., 1991. Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models. Econometrica 59, 1551–1580.

Johansen, S., Juselius, K., 1990. Maximum likelihood estimation and inference on cointegration – with applications to the demand for money. Oxford Bulletin of Economics and Statistics 52, 169–210.

Jorion, P., Sweeney, R., 1996. Mean reversion in real exchange rates: evidence and implications for forecasting. Journal of International Money and Finance 15, 535–550.

Kim, Y., 1990. Purchasing power parity in the long run: a cointegration approach. Journal of Money, Credit and Banking 22, 348–357.

Keynes, J.M., 1932. Essays in Persuasion, Harcourt Brace, New York.

Kremers, J.J.M., Ericsson, N.R., Dolado, J.J., 1992. The power of cointegration tests. Oxford Bullettin of Economics and Statistics 54, 325–348.

Lothian, J.R., 1990. A century plus of yen exchange rate behavior. Japan and the World Economy 2, 47–90.

Lothian, J.R., Taylor, M.P., 1996. Real exchange rate behavior: the recent float from the perspective of the past two centuries. Journal of Political Economy 104, 488–510.

Lothian, J.R., Taylor, M.P., 1997. Real exchange rate behavior: the problem of power and sample size. Journal of International Money and Finance 16, 945–954.

Lucas, R.E., 1982. Interest rates and prices in a two currency world. Journal of Monetary Economics 10, 335–360.

MacKinnon, J.G., 1991. Critical values for cointegration tests. In: Engle, R.F., Granger, C.W.J. (Eds.), Long Run Economic Relationships: Readings in Cointegration. Oxford University Press, Oxford.

McKinnon, R.I., 1971. Monetary Theory and Controlled Flexibility in the Foreign Exchanges. International Finance Section, Princeton University, Princeton, NJ.

Mark, N., 1990. Real and nominal exchange rates in the long run: an empirical investigation. Journal of International Economics 28, 115–136.

McNown, R., Wallace, M., 1989. National price levels, purchasing power parity and cointegration: a test of four high inflation economies. Journal of International Money and Finance 8, 533–545.

Mussa, M., 1986. Nominal exchange rate regimes and the behavior of real exchange rates: evidence and implications. Carnegie-Rochester Conference Series on Public Policy 25, 177–213.

Obstfeld, M., Rogoff, K., 1996. Foundations of International Macroeconomics. MIT Press, Cambridge, MA.

Obstfeld, M., Rogoff, K., 1995. Exchange rate dynamics redux. Journal of Political Economy 103, 624–660.

O'Connell, P.G.J., 1998. The overvaluation of purchasing power parity. Journal of International Economics 44, 1–19.

Officer, L.H., 1982. Purchasing Power Parity and Exchange Rates: Theory, Evidence and Relevance. JAI Press, Greenwich, CT.

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

Papell, D.H., 1998. Searching for Stationarity: Purchasing Power Parity Under the Current Float. Journal of International Economics 43, 313–332.

Piggott, C., Sweeney, R.J., 1985. Purchasing power parity and exchange rate dynamics. In: Arndt, S.W., Sweeney, R.J., Willett, D. (Eds.), Exchange Rates, Trade and the US Economy. Ballinger and American Enterprise Institute, Cambridge, MA.

Ramsey, J.B., 1969. Tests for specification errors in classical linear least squares regression analysis. Journal of the Royal Statistical Society, series B, 31, 350–371.

Reinsel, G.C., Ahn, S.K., 1988. Asymptotic Properties of the Likelihood Ratio Test for Cointegration in the Non Stationary Vector AR Model. Technical Report, Department of Statistics, University of Wisconsin, Madison.

Rogoff, R., 1996. The purchasing power parity puzzle. Journal of Economic Literature 34, 647–668.

Roll, R., 1979. Violations of purchasing power parity and their implications for efficient international commodity markets. In: Sarnat, M., Szego, G.P. (Eds.), International Finance and Trade, vol 1. Ballinger, Cambridge, MA.

Samuelson, P.A., 1964. Theoretical notes on trade problems. Review of Economics and Statistics 46, 145–154.

Shiller, R.J., Perron, P., 1985. Testing the random walk hypothesis: power versus frequency of observation. Economics Letters 18, 381–386.

Schwartz, G., 1978. Estimating the dimension of a model. The Annals of Statistics 6, 461–464.

Stockman, A.C., 1980. A theory of exchange rate determination. Journal of Political Economy 88, 673–698.

Taylor, M.P., 1988. An empirical examination of long run purchasing power parity using cointegration techniques. Applied Economics 20, 1369–1381.

Taylor, M.P., 1995. The economics of exchange rates. Journal of Economic Literature 33, 13–47.

Taylor, M.P., McMahon, P.C., 1988. Long run purchasing power parity in the 1920s. European Economic Review 32, 179–197.

Taylor, M.P., Sarno, L., 1998. The Economics of Exchange Rates. Cambridge University Press, Cambridge and New York, forthcoming.

Wei, S., Parsley, D., 1995. Purchasing Power Disparity During the Floating Rate Period: Exchange Rate Volatility, Trade Barriers and Other Culprits. National Bureau of Economic Research, Working paper 5032.

West, K.D., 1988. Asymptotic normality when regressors have a unit root. Econometrica 56, 1397–1417.

Wu, Y., 1996. Are real exchange rates non stationary? Evidence from a panel data test. Journal of Money, Credit and Banking 28, 54–63.

Zellner, A., 1962. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American Statistical Association 57, 348–368.