No. 2120

HOW SURE ARE WE ABOUT PPP? PANEL EVIDENCE WITH THE NULL OF STATIONARY REAL EXCHANGE RATES

Biing-Shen Kuo and Anne Mikkola

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

ISSN 0265-8003

HOW SURE ARE WE ABOUT PPP? PANEL EVIDENCE WITH THE NULL OF STATIONARY REAL EXCHANGE RATES

Biing-Shen Kuo and Anne Mikkola

Discussion Paper No. 2120
March 1999

Centre for Economic Policy Research
90–98 Goswell Rd
London EC1V 7RR
Tel: (44 171) 878 2900
Fax: (44 171) 878 2999
Email: cepr@cepr.org

CEPR Discussion Paper No. 2120

March 1999

ABSTRACT

How Sure Are We About PPP?
Panel Evidence with the Null of Stationary Real Exchange Rates*

There has been serious suspicion of a spurious rejection of the unit roots in panel studies of purchasing power parity (PPP) due to the failure to control for cross-sectional dependence. This article presents evidence of mean-reversion in industrial country real exchange rates in a set-up that accounts naturally for cross-sectional dependence, is invariant to the benchmark currency and capable of detecting against regime changes, and actually tests for the null of interest, i.e. the purchasing power parity. Our results are based on a Kwiatkowski, Phillips, Schmidt and Shin (KPSS, 1992) test for the stationarity null generalized in a multivariate random walk plus noise model by Nyblom and Harvey (1998).

JEL Classification: C22, C52, F31, F47
Keywords: real exchange rates, unit root

Biing-Shen Kuo
Graduate Institute of International Trade
National Chengchi University
Taipei 116
TAIWAN
Tel: (886 2) 939 3091
Fax: (886 2) 937 9071
Email: bsku@nccu.edu.tw

Anne Mikkola
Research Unit on Economic Structures and Growth
Department of Economics
PO Box 54
University of Helsinki
Helsinki
FINLAND
Tel: (358 9) 191 7984
Fax: (358 9) 191 7980
Email: anne.mikkola@helsinki.fi

*This work is a revised version of a paper entitled ‘Testing for PPP in a Panel Industrial Countries Allowing for Cross-Sectional Dependence’ which was presented at the 8th International Conference for Panel Data, Göteberg, Sweden, June 1998, at the ESEM and at the EEA meeting, Berlin, Germany, September 1998. The authors wish to thank the conference participants for their valuable comments. They also gratefully acknowledge the Academy of Finland and the Yrjö Jahnsson Foundation grants to the Research Unit on

Economic Structures and Growth (RUESG) and the National Science Council grants (NSC–88–2415–H–004–010).

Submitted 15 January 1999

NON-TECHNICAL SUMMARY

The question we are asking is the following: Are some movements in real exchange rates permanent or is there an equilibrium real exchange rate towards which the exchange rate between any two countries (deflated by their respective price levels) eventually returns? In other words, we want to assess the usefulness of the purchasing power parity (PPP) theory as a first approximation to the long-run behaviour of exchange rates. To tackle the question, we want to make use of as much data as possible by using a panel of industrial country data. The main differences of our study from the previous ones is that we apply testing methods that take into account the dependence between currencies and test for the actual hypothesis, the PPP, as the null. These are issues that have been shown to be critical for the results in the panel literature until now. Econometric difficulties have posed problems in tackling them, however.

Why then is the question of permanence of shocks to real exchange rates (i.e. existence of a unit root) versus their stationarity such an important question? For purchasing power parity to hold in the long run, the real exchange rates must be stationary. Permanent shocks to them would imply a permanent tendency for the purchasing powers of the currencies to deviate from one another. Whether real exchange rates are stationary or non-stationary matters, since the two alternatives are associated with two quite different long-run economic implications. In the former case, PPP serves as a good first approximation to the long-run behaviour. This is the view of practising economists when they base their long-run exchange rate forecasts on some measures of equilibrium real exchange rates, or make decisions on fixing parities between currencies. In the latter case, PPP serves no purpose, even over the long run. The finding of a unit root in the real exchange rates would be problematic for many of the established theories. Furthermore, it would make long-run forecasting a useless exercise.

Having said all this, do we mean we believe in a possibility of an 'eternal' equilibrium real exchange rate? Few economists would go this far. There is evidence that the price levels in rich countries tend to be higher than in poor countries when converted to a common currency. The evidence for the industrial countries is more debatable, however. Thus, what we are really testing empirically is if the permanent deviations from PPP are of relatively minor importance. In that sense, the long-run PPP and its usefulness, is an empirical matter that can be tackled by testing for stationarity of the real exchange rates.

Empirical testing of PPP has predominantly focused on testing for whether the unit root can be rejected in the real exchange rates. These tests on relatively short individual time series, e.g. on post-1973 data, typically fail to support PPP. The emerging ‘consensus’ of the failure of PPP started to shift back towards acceptance of the PPP in the 1990s, however. Studies using longer time series were able to reject the unit root. A similar message supporting PPP was given by the studies making use of data from a panel of countries. Somewhat remarkably, they reported similar speeds of adjustment to PPP as the studies based on a single country data – half-life of deviations from PPP being in the range of 3–5 years.

Recently, doubts have been raised that the emerging support for PPP in the panel studies may be due to the failure to account for dependence between the currencies. Real exchange rates are known to be highly dependent, not only because of economic dependence between the countries’ price levels and exchange rates, but also because of the way they are constructed by calculating them with respect to common benchmark currency.

The aim of this paper is to judge the extent to which we can think of PPP holding, given that the dependence in the real exchange rates may have seriously biased the inference. We also want to guard against rejecting the theory too lightly. It seems unwarranted to draw quick conclusions on PPP not holding based on not being able to reject the unit root – the null that is not our hypothesis. Simulation results indeed illustrate how uninformative the rejection of the null of panel unit root tests can be about PPP.

We apply a recently developed test that appears ideal for our purposes. It has an advantage of incorporating dependence between the real exchange rates very naturally in the estimation. Unlike the most frequently used panel version of the unit root test, our test is a test for stationarity. Since PPP is our hypothesis, we indeed ought to be having stationarity of the real exchange rates as the null. An additional feature of the test is its invariance to the choice of benchmark currency. This is a desirable property, given that empirical studies tend to report more support for PPP when the Deutsche mark rather than the US dollar is used as a benchmark currency.

The test will be applied to a panel of industrial country real exchange rates over 1949–96. It may be reasonable to focus on a relatively homogeneous group of countries when using panel methods. Also, it does not seem reasonable to expect PPP to hold between countries at very different levels of development. It has been argued that the use of data from pre- and post-Bretton Woods period is not appropriate in PPP studies because of regime changes. Consequently, much of the empirical work has looked at the floating rate regime after 1973 data only. We think, however, that as a long-run phenomenon, PPP should apply regardless of the exchange rate regime.

The purchasing power parity hypothesis cannot be rejected for our panel of 23 industrial countries, nor for the subsample of 12 European community countries over 1949–96. PPP is found to hold for the European countries on the post-Bretton Woods period as well. Additional complementary evidence in favour of PPP comes from applying the panel version of the conventional unit root test. Our results indicate that the real exchange rates of these countries tend to return to the long-run equilibrium, even if there is a great deal of short-term variation. The doubts that there is no support for PPP when the cross-sectional dependence is accounted for, is therefore not supported by our study. In contrast to most recent panel studies, our evidence is based on cross-country data spanning both the current float and the previous regime. The results reinforce the view that PPP, as a long run phenomenon, has little to do with the exchange rate regimes. The evidence presented in the paper, as in the recent literature, reveals that purchasing power parity remains to serve us well as a first long-run approximation. The PPP hypothesis seems to find its voice echoing after all.

1 Introduction

For purchasing power parity (PPP) to hold in the long run, the real exchange rates must be stationary. Permanent shocks to them would imply a permanent tendency for the purchasing powers of the currencies to deviate from one another. Whether real exchange rates are stationary or non-stationary matters, since the two alternatives are associated with two quite different long run economic implications. In the former case, PPP serves as a good first approximation to the long-run behavior. This is the view of the practising economists when they base their long-run exchange rate forecasts on some measures of equilibrium real exchange rates, or make decisions on fixing parities between currencies. In the latter case, PPP serves no purpose, even over the long-run. The finding of a unit root in the real exchange rates would be problematic for many of the established theories. Furthermore, it would make long run forecasting a useless exercise.

Having said all this, do we mean we believe in a possibility of an `eternal' equilibrium real exchange rate? Few economists would go this far. There is evidence that the price levels in rich countries tend to be higher than in poor countries when converted to a common currency, due to e.g. the Balassa-Samuelson effect. The evidence for the industrial countries is more debatable, however (see Rogoff, 1996). Thus, what we are really testing empirically is if the permanent deviations from PPP are of relatively minor importance. In that sense, the long-run PPP, and its usefulness, is an empirical matter that can be tackled by testing for stationarity of the real exchange rates.

Empirical testing of PPP has predominantly focused on testing for whether the unit root can be rejected in the real exchange rates. Unit root or cointegration tests on relatively short univariate time series, e.g. on post-1973 data, typically fail to support PPP. The emerging `consensus' of the failure of PPP started to shift back towards acceptance of the PPP in the 90's, however. Studies using longer time series were able to reject the unit root. A similar message supporting PPP was given by the panel studies. Somewhat remarkably, they reported similar speeds of adjustment to PPP as the long sample studies, half live of deviations from PPP being in the range of 3-5 years.[1]

Recently, O'Connell (1998) has put forth evidence questioning the support for PPP found in

[1] For discussion of the empirical results other than the very latest ones see Froot and Rogoff (1995) for an excellent survey. The success of panel studies (e.g. Coakley and Fuertes, 1997; MacDonald, 1996; Oh, 1996; Wei and Parsley, 1995; Wu, 1996) to find support for PPP is considered to be due to increased power of the unit root tests with more observations and more variation in data.

panel studies. His simulations show that the rejection of the unit root in the panel studies may be entirely due to the failure to account for cross-sectional dependence. Real exchange rates are known to be highly dependent, not only because of economic dependence between the countries' price levels and exchange rates but also due to the construction of the real exchange rates with respect to some common benchmark. Yet, this correlation has been largely ignored because of econometric difficulties in dealing with it. Typically, the panel version of the standard univariate unit root test (Levin and Lin, 1993; revised as Levin, Lin and Chu, 1997, henceforth LLC[2]) is applied. The LLC-test has been particularly attractive for testing PPP, since it made possible the use of more data, alleviating the main problem in econometric inference. Thus, the literature on testing PPP grew, even if it made the obviously unrealistic assumption of no cross-country dependence. There were few studies making an attempt to tackle the problem.

The aim of this paper is to judge the extent to which we can think of PPP holding, given the fact that the dependence in the real exchange rates may have seriously biased the inference in the panel studies. We also want to guard against rejecting the theory too lightly. It seems unwarranted to draw quick conclusions on PPP not holding based on not being able to reject the unit root - the null that is not our hypothesis. On the other hand, the Monte Carlo simulations by Taylor and Sarno (1997) illustrate, in an eye-opening manner, how uninformative the rejection of the null of panel unit root tests can be about PPP. Caution in interpretation of these tests is in order: the null hypothesis that all the series are realizations of unit root processes will be violated even if only one of the series is stationary. The simulations show that even with only one stationary and persistent process in the panel, the unit root tests may lead to a very high probability of rejection of the null. We would not want to confidently claim that PPP holds in this case - even if we find the null being rejected.

An ideal test for our purposes has been recently developed by Nyblom and Harvey (1998, henceforth NH). This test has a major advantage of incorporating cross-sectional dependence very naturally in the panel estimation. Unlike the standard panel version of the ADF-test, the NH test is a test for the null of stationarity. Since PPP is our hypothesis, we indeed ought to be having

[2] The revision is concerned with the proofs of the asymptotic results. It does not affect the conduct of the test or the critical values derived in the earlier version.

stationarity of the real exchange rates as the null. An additional feature of the NH-test is its invariance to the choice of benchmark currency. This is a desirable property, given that empirical studies tend to report more support for PPP when German mark rather than the US dollar is used as a benchmark currency. We also contrast our results, with and without cross-sectional dependence, to those of the tests with the null of unit root.

The test will be applied to a panel of industrial country real exchange rates over 1949-1996. It may be reasonable to focus on a relatively homogeneous group of countries when using panel methods. Also, it does not seem reasonable to expect PPP to hold between countries at very different levels of development. Industrial countries are a category in many previous panel studies, thus allowing for comparison of the results. It has been argued that the use of data from pre- and post-Bretton Woods period is not appropriate in PPP studies because of regime changes. Consequently, much of the empirical work has looked at the floating rate regime after 1973 data only. We, however, think that as a long run phenomena PPP should apply regardless of the exchange rate regime. Some empirical support for this view can be seen in Lothian and Taylor (1996). They show that the stationary process estimated for pre-73 data performs well in out-of-sample forecasting for the post-73 period as well. Froot, Kim and Rogoff (1995) also point to the irrelevance of the exchange rate regime over the long run: There is evidence on the surprising stability in the volatility and persistence of deviations from the law of one price over 700 years in England and Holland. Our results also lend support to this view. The NH test is indeed capable of discovering the alternatives with structural shifts, besides the alternatives with nonstationarity. The values of the actual sample statistics over pre- and post-Bretton Woods periods, based on our simulations, appear to come rather from a stable stationary than from a regime-dependent non-stationary process. We will, however, do our testing exercise for the post Bretton Woods period as well when possible. Also the subset of European countries are considered.

The paper is built as follows. Section 2 discusses the standard setup for testing PPP in panels, and the sources and the treatment of cross-sectional dependence. The Nyblom-Harvey test allowing for cross-sectional dependence and its properties are discussed in section 3. Section 4 contains our results. Section 5 compares the results to the previous literature with unit root null. Section 6 concludes.

2 Standard testing for PPP in panels: the null of unit root

Typically, the real exchange rates are constructed from the consumer price index series and the exchange rate series for the price of U.S. dollars in respective currencies. Real exchange rate for country i at time t is thus

$$q_{it} = p_{it} - e_{it} - p_{USt};$$

where p_{it} is the CPI for country i at time t, p_{USt} is the CPI for the US and e_{it} is the exchange rate of country i at time t in units of country i's currency per US dollar. All variables are in logarithms.

In the panel set up, all individual real exchange rates are calculated. The bilateral PPP relationships are stacked and the panel methods are applied. The most frequently used test is the test of Levin-Lin-Chu (1997), which is particularly suited for moderately sized panels, such as encountered in cross-country studies. It is attractive, because it allows for incorporation of different individual-specific effects under the null. PPP can be tested in a panel of N countries by testing if $\rho < 0$ in the following ADF regressions

$$\Delta q_{it} = \rho q_{i,t-1} + \sum_{L=1}^{p} \phi_{iL} \Delta q_{it-L} + \alpha_i + \varepsilon_{it} \qquad \text{for } i = 1, \ldots, N \tag{1}$$

where α_i's are country specific intercepts. ε_{it}'s are assumed to be uncorrelated across countries and time.

Under the null, $\rho = 0$ and $\alpha_i = 0$ for all i, there exists a unit root in all the N real exchange rates, and PPP does not hold. In other words, deviations from PPP in each series tend to be lasting. The economic reasoning behind permanent deviations can be understood, for example, in an equilibrium model of exchange rates as outlined in Stockman (1987). The real exchange rate in that model, equal to the marginal rate of substitution between home and foreign goods, would exhibit a random walk in response to a permanent shift in demand preference.

Under the alternative, as a result of arbitraging flow of goods and services, some of these deviations would tend to die off, but not necessarily all of them. This is because the null is a composite hypothesis that all the N real rates are non-stationary, and the existence of at least one stationary real rate series in the panel would suggest a violation of the null. This is one of potential shortcomings with using LLC test in the previous studies. A rejection of the null by the test is not sufficiently informative about whether all the series under consideration are stationary. Taylor and Sarno (1997) emphasize that a caution in interpreting a rejection by the test needs to be exercised.

Levin el al. (1997) showed that the t-value of $\hat{\rho}$ in equation (1) under the null has a limiting distribution that is normal. [3] This result is, however, dependent on the assumption that is very inappropriate in real data, i.e. the cross-sectionally uncorrelated regression disturbances in eq. (1).

The consequences of ignoring the cross-sectional dependence can be dramatic. O'Connell (1998) demonstrates this in his Monte Carlo simulations. Significance levels of tests with a nominal size of 5 percent rise to as much as 50 percent. This means that the common panel tests tend to reject the null of unit root far too frequently, giving false support for PPP. The favorable findings to PPP by the previous panel studies, thus, might simply be the reflection of a spurious rejection of the unit root by the test.[4] The power of the tests is adversely affected by the cross-sectional dependence as well.

Cross-sectional dependence may arise for two reasons. Firstly, the real exchange rates are correlated by construction. They all contain a common component arising through the common benchmark, since the real exchange rates are calculated with reference to a single benchmark country, typically the US. Consequently, any independent variation in the benchmark country's price level or the value of its currency shows in all the real exchange rates.

Secondly, the price indices and the exchange rates that underlie the constructed real exchange rates are not independent ie. there is economic dependence between the countries. How does the economic dependence between countries transmit into PPP relationships? Real supply or demand shocks can be common to all or a subset of countries, or they can be country specific shocks. In either case, the terms of trade must adjust, and this requires movements in exchange rates and prices to a different extent in different countries. The interdependencies of economic quantities differ across countries due to different type of trade ties between countries. All this is bound to

[3] This is in contrast to the t-value of $\hat{\rho}$ in individual time series, which has a non-normal limiting distribution.

[4] These may include, among others, Frankel and Rose (1996), Oh (1996), Papell (1997), Wei and Parsley (1995) and Wu (1996).

make real exchange rates correlated but to a different degree across different countries. See e.g. Canova and Dellas (1993) and Stockman and Tesar (1995) for discussion on the economic linkages.

Can cross-sectional dependence be accounted for in panel unit root testing? As suggested by Levin et al. (1997) and Im et al. (1995), cross-sectional dependence can be accounted for partially by removing the time specific effect by time dummies. This amounts to subtracting the cross-section averages from the data, thus removing the common component. The limiting distributions of the panel unit root test statistics are not affected. We can then think of the error term consisting of a stationary time specific common effect and an idiosyncratic random effect. This procedure is appropriate when there is a single aggregate common factor, which has an identical impact on all individuals in the panel. In our case the fluctuation of the US dollar clearly is such a common factor for the individual dollar denominated real exchange rates.

There are few studies besides O'Connell's paying attention to cross-sectional dependence arising due to economic dependence. Hakkio (1984), Abuaf and Jorion (1990), Jorion and Sweeney (1996) and Taylor and Sarno (1997) all find support for PPP. Common to all these studies, including O'Connell's, is that they resort to bootstrapping to derive the empirical distribution of the unit root test statistic under cross-sectional dependence. This is necessary since the distribution of the test statistics like Levin and Lin's is not invariant to the covariance matrix of the error terms. To obtain the correct critical values in each case, an estimate for the error covariance matrix is needed. And thus critical values need to be simulated separately for every case.

3 Testing for PPP allowing for cross-sectional dependence: the null of stationarity

In this section, we advocate an approach that deals directly with the issues with interpretation of test and cross-sectional correlation in the extant panel studies, while avoiding the use of the LLC test. The recent test developed by Nyblom and Harvey (1998) renders this feasible. It is a multivariate extension of the Kwiatkowski, Phillips, Schmidt and Shin (KPSS, 1992) test. For illustration, we present the model first in its simplest form with no serial correlation. The panel of real exchange rate series forms a multivariate model of random walk plus noise,

$$q_t = \delta + \mu_t + \varepsilon_t; \quad \varepsilon_t \sim iid(0, \Sigma_\varepsilon)$$

$$\mu_t = \mu_{t-1} + \eta_t; \quad \eta_t \sim iid(0, \Sigma_\eta)$$

where

$$q_t = (q_{1t}, q_{2t}, \ldots, q_{Nt})', \; \delta_t = (\delta_{1t}, \delta_{2t}, \ldots, \delta_{Nt})', \text{ and } \mu_t = (\mu_{1t}, \mu_{2t}, \ldots, \mu_{Nt})',$$

where q_t is the vector containing N real exchange rate time series, δ is the non-zero intercept vector whose elements vary with countries but not with time, and μ_t is a random-walk time series. The test statistics is

$$\xi_N = tr[S^{-1}C], \tag{2}$$

where S is an estimate for the so-called long-run error covariance matrix computed as

$$S = T^{-1} \sum_{t=1}^{T} (q_t - \bar{q})(q_t - \bar{q})', \text{ where } \bar{q} = \frac{1}{T} \sum_{t=1}^{T} q_t,$$

and

$$C = T^{-2} \sum_{j=1}^{T} \left[\sum_{t=1}^{j} (q_t - \bar{q}) \right] \left[\sum_{t=1}^{j} (q_t - \bar{q}) \right]'.$$

3.1 Stationarity as the null

As in the KPSS test, ξ_N is a test for stationarity under the null. Testing for stationarity corresponds to testing the hypothesis that there is no random-walk component in the system. Equivalently, under

$$H_0: \Sigma_\eta = 0,$$

all real exchange rates are stationary. Under the alternative, all the real exchange rates, or part of them, have unit roots.

Having stationarity of real exchange rates as the null, rather than the alternative, is desirable because of the difficulty of the panel unit root tests to decide whether PPP gets supports. This testing strategy is also consistent with the way the classical hypothesis testing is performed. It ensures that the null of PPP is not rejected as long as there is no strong evidence against it.

Stationarity will be rejected for large values of ξ_N. The limiting distribution of ξ_N under the null is a Cramer-von Mises distribution with N degrees of freedom, $CVM(N)$:

$$\xi_N \xrightarrow{d} \int_0^1 B(u)'B(u)du = \sum_{k=1}^{\infty} (\pi k)^{-2} u_k' u_k; \qquad (3)$$

where u_k is an $N \times 1$ vector such that $u_k \sim NID(0; I_N)$ and $B(u)$ is a standard vector Brownian bridge of dimension N. Notice that the asymptotic distribution depends only on the number of time series, N. Nyblom and Harvey tabulate the critical values only for up to $N = 4$. We will thus need to compute the critical values for N greater than 4 separately.[5]

It is natural to ask what the advantage of the NH test over the KPSS test is when applying both to a panel. To answer this, we note that the partial residual sum is the basic ingredient in constructing the tests. For a panel of real rate series, the KPSS test looks only at the partial sum square of each series (the diagonal elements in C, $[\sum_{t=1}^{T}(q_{kt} - \bar{q}_k)]^2$), while the NH test takes into account not only that but also correlations of the partial sums of different series (the off-diagonal elements in C, $[\sum_{t=1}^{T}(q_{kt} - \bar{q}_k)][\sum_{t=1}^{T}(q_{lt} - \bar{q}_l)]$). This shall enable the NH test to have more opportunities in detecting stationarity property in the data than the KPSS test.[6] This is also a merit comparable to a greater precision in estimating mean-reverting parameters associated with the LLC test (ρ in eq.(1)) due to more variation with panel data.

3.2 Allowing for cross-sectional dependence

Another important advantage of the test is that it accounts for cross-sectional dependence naturally. In notation, this implies that the error covariance matrix, Σ_ε, needs not be diagonal. This practically allows the testing procedure to capture information regarding correlations among real rates, and thus has an implication to the test power. When there is no cross-sectional dependence, ie, Σ_ε is diagonal, all errors are mutually independent over time and countries.

It should be emphasized that the asymptotic distribution of the test statistic is invariant to the error covariance matrix: No bootstrapping is necessary and the sensitivity of inference to different

[5] This can be done following the formula in (3) with a sufficiently large truncation number to approximate the distribution well enough.

[6] Choi and Ahn (1999) mount tests for the stationarity null for multiple time series. One of them, signified as SBDH, takes the same form as the NH test. Their simulations demonstrate that the SBDH test is more powerful than the univariate counterparts.

covariance matrix estimators can be directly tested since the critical values stay fixed.

3.3 Invariance property

The NH test is also featured by its invariance to the choice of the benchmark currency.[7] Many studies report stronger support for PPP (rejections of the unit root), when the German mark is used as a numeraire rather than the US dollar. This may be related to the large upward and the consequent large downward swing of the dollar in the 80's that dominates all the dollar denominated real exchange rates.[8] This "numeraire effect" seems unjustifiable on theoretical grounds. By construction, a new panel of real exchange rates based on a different currency is simply a linear transformation from one panel based on another currency. Thus, if one panel is stationary, another would be so as well. In this sense, the NH test is desirable as a test of PPP, because of its irrelevance to the base currency under the stationarity hypothesis.

3.4 Detecting against regime changes

Our testing strategy makes it possible to examine PPP using panels that encompass fixed and floating regimes as long as available, at the same time be able to detect if the behavior of real exchange rates varies with the regimes. Rationale for avoiding the use of data with long time span in the PPP studies has been that they are possibly subject to structural changes (e.g. Frankel and Rose, 1996). Applying the NH test, however, could escape the concern, because it in fact possesses good power against regime changes. The random-walk component under the alternative of the test indeed represents a random variation from the mean of the model. By regarding a structural break as a variation that constantly occurs, the test would be capable of detecting the break when it

[7] Constructing a new panel of real exchange rates by benchmarking against a different country currency than USD can be achieved by the following transformation,

$$\tilde{q}_t = A q_t$$

where A is the non-singular transformation matrix as defined in the appendix of O'Connell (1998). Letting $\tilde{S} = ASA' = T^{-1}\sum_{t=1}^{T}(\tilde{q}_t - \bar{\tilde{q}})(\tilde{q}_t - \bar{\tilde{q}})'$, $\tilde{C} = ACA' = T^{-2}\sum_{j=1}^{T}[\sum_{t=1}^{j}(\tilde{q}_t - \bar{\tilde{q}})][\sum_{t=1}^{j}(\tilde{q}_t - \bar{\tilde{q}})]'$, where C and S are defined in the text, the invariance property holds simply because

$$\xi_N = tr[S^{-1}C] = tr[(A'^{-1}S^{-1}A^{-1})(ACA')] = tr[\tilde{S}^{-1}\tilde{C}].$$

[8] For discussion on the special role of the US dollar in the PPP studies, see e.g. Lothian (1998), Papell and Theodoridis (1998) and Koedijk et al. (1998).

exists in the data.[9] The results for the test with regime changes are discussed in the next section.

3.5 The econometric issues

For the test to be useful in practice, we need to allow for serial correlation. The real exchange rate for country i at time t is written now as

$$\alpha_i(L)q_{it} = \delta_i + \mu_{it} + \varepsilon_{it}; \qquad (4)$$

where $\alpha_i(L) = 1 - \alpha_{i1}L - \cdots - \alpha_{ip_i}L^{p_i}$ is a autoregressive polynomial in the lag operator with the root outside the unit circle. Note that the lag orders with each real rate (p_i) may differ from each other. Restriction to this sort of models allows for reasonable flexibility. For our particular sample, even a model with two lags is found to fit the data well. The model also allows for various convergence speeds across countries as no restriction is placed on the AR coefficients.

As suggested by Nyblom and Harvey, a nonparametric adjustment for serial correlation along the line of KPSS (1992) can be applied. This means replacing S above by

$$S(m) = \hat{\Gamma}_0 + \sum_{\tau=1}^{m} \omega(\tau)[\hat{\Gamma}_\tau + \hat{\Gamma}'_\tau]$$

where $\omega(\tau)$ is a weighting function or a kernel with $\omega(\tau) = 1 - \tau/(m+1); \tau = 1; \ldots; m$ (the Bartlett kernel) and

$$\hat{\Gamma}(\tau) = T^{-1}\sum_{t=\tau+1}^{T}(q_t - \bar{q})(q_{t-\tau} - \bar{q})'$$

is the sample autocovariance matrix at lag τ. This is a way of estimating the long run covariance matrix of the noise process, and thus getting an estimate for the cross-sectional dependence. It is known that the estimated covariance matrix can be very sensitive to the choice of the bandwidth parameter (m) in the case of the non-parametric kernel based estimators.[10] The sensitivity of the results to different lag truncation points (m) in the Bartlett kernel will be checked and reported.

[9] Kuo (1998) employs this idea in constructing tests for constancy in cointegrating coefficients. His L_c test inherits the same structure as the NH test, and is found to be quite powerful against both random and permanent shifts on the coefficients.

[10] For the non-parametric estimators, the choice of the bandwidth parameter is particularly problematic, since there is no way of choosing an optimal value of the parameter for any particular finite sample. The optimality criteria used only gives the rate at which the bandwidth parameter should grow as a function of sample size. The true covariance matrix would weigh all the sample autocovariances equally. Kernels with declining weights are, however, used to ensure positive semi-definite covariance matrix. This leads to a trade-off between the bias and the variance of the estimated covariance matrix. The larger the bandwidth, the less the bias incurred by placing weights less than one on autocovariances at lags shorter than the sample length. At the same time, raising the bandwidth puts a larger weight on the higher sample autocovariances that are relatively poorly estimated.

Alternatively, we apply a parametric method of estimating the long run covariance matrix. Parametric estimation has the advantage that model selection criteria can be used to evaluate the trade-off between goodness of fit and parsimony (see Den Haan and Levin, 1996 for discussion of different long-run covariance matrix estimators). We apply the NH test to the panel of real exchange rates data from which serial correlation is first being removed. The idea is presented in Leybourne and McCabe (1994). For the NH test to have power, the AR parameters need to be estimated in a way that they are consistent under both the null and the alternative. For a univariate case, the model in (4) after differencing can be written as follows.

$$\alpha_i(L)\Delta q_{it} = (1 - \theta_i L)\zeta_{it}; \qquad (5)$$

where $\alpha_i(L)$ is as defined before, and $\sigma^2_{\zeta_i}$ is distributed as $iid(0, \sigma^2_{\zeta_i})$.[11]

The model in (5) can be understood as a reduced form of the structural model in (4). Testing for the stationarity is now equivalent to testing if $\theta_i = 1$. Thus, under $H_0 : \theta_i = 1$, the model (5) is a stationary $AR(p_i)$ process, and under $H_a : |\theta_i| < 1$ it is a nonstationary $ARIMA(p_i, 1, 1)$ process.

To get consistent estimates of $\alpha_i(L)$ under the null and the alternative, Leybourne and McCabe (1994) suggest fitting an $ARIMA(p_i, 1, 1)$ to each of the real exchange rate series as in (5). BIC (Bayes Information Criterion) is used to select the AR order for each country individually. The residual series accounting for serial correlation,

$$q^*_{it} = q_{it} - \hat{\alpha}_i(L) q_{it}$$

are formed, and the NH test is applied to the panel consisting of them, $q^*_t = (q^*_{1t}, q^*_{2t}, \ldots, q^*_{Nt})'$.

The asymptotic distribution of the NH-test statistics remains the same, $CVM(N)$, regardless of the estimator of the long run covariance matrix. This is a major advantage relative to the previous methods of deriving bootstrap distributions for each estimated covariance structure individually. In the Nyblom-Harvey set up, we can thus easily check for the sensitivity of the inference to the alternative estimates of the covariance between the real exchange rates.

[11] ζ_{it} is a function of ε_{it} and η_{it}. Their variances are linked as follows. Defining $r_i = \sigma^2_{\eta_i}/\sigma^2_{\varepsilon_i}$, θ_i is then related to r_i by $r_i = (1 - \theta_i)^2/\theta_i$. We notice that $r_i = 0$ if and only if $\theta_i = 1$.

4 Data and empirical results

4.1 Preliminary data analysis

Data is extracted from the IMF publication, International Financial Statistics. The consumer price index series (IFS line 64) is used as the measure of prices, and the price of U.S. dollars in respective currencies (IFS line rf) as the exchange rate. A panel of annual real exchange rates is then constructed. Data is available for the following 24 countries from 1949-1996: the US, the UK, Austria, Belgium, Denmark, France, Germany, Italy, Luxembourg, Netherlands, Norway, Sweden, Switzerland, Canada, Japan, Finland, Greece, Ireland, Malta, Portugal, Spain, Australia, New Zealand and South Africa. We apply the NH test to the panel as a whole, and two subpanels: the 12 European countries over the entire sample period and over the post 1973 period, respectively.[12] Annual data is used, since we are looking at the long-run behavior rather than short-run fluctuations. It has been pointed out, that the power of the unit root tests depends more on the span of the data in years than on the number of observations per se (see Shiller and Perron, 1985).

It is important to verify the significance of cross-sectional dependence first. To get an estimate of the innovation covariance matrix of the real exchange rates, SURE estimation is applied to the panel consisting of the 23 AR(p) models - one for each real exchange rate. The best fitting AR order for each real exchange rate series is chosen by the Bayes information criteria before the SURE estimation. Likelihood ratio (LR) test is then employed to test if the off-diagonal elements in the innovation covariance matrix can be restricted to zero.[13] Indeed, the values of the χ^2 $-$statistics show clear evidence against no cross-sectional dependence in all the panels under study.

4.2 Results

The values of the NH ξ_N -statistic are reported in Panel A of Table 1 for the full panel of 23 industrial countries and the 12 European countries with the non-parametric and parametric estimators of the long-run covariance matrix. The results for the full sample period 1949-96 and the post Bretton

[12] The EC12 panel includes the UK, Belgium, Denmark, France, Germany, Italy, Luxembourg, Netherlands, Greece, Ireland, Portugal and Spain.

[13] The test is computed as $T\{\ln|\hat{\Sigma}(0)| - \ln|\hat{\Sigma}|\}$. $\hat{\Sigma}(0)$ and $\hat{\Sigma}$ are the estimated covariance matrices under the null with no cross-sectional dependence and under the alternative with no restriction imposed respectively. T is the time length. The test has an asymptotic $\chi^2(66)$ $-$ distribution in the panel of 12 EC countries and a $\chi^2(253)$ $-$ distribution in the in full dataset of 23 real exchange rates. The value of the test-statistic is 1800 in the full dataset. In the EC dataset for the full sample period and post-1973 period the values are 1010 and 697 respectively.

Woods period (1973-96) are reported for the EC countries.[14] The asymptotic and small sample critical values for the NH test for our panel dimensions are calculated in Panel B of Table 1.

The results emerging from these tables are very similar for all the panels. The null of stationary real exchange rates cannot be rejected for any of the panels, typically not even at 10% level of significance, when the small sample critical values are used. This result is robust to the choice of the bandwidth parameter (m=4,8 or 12) and the methods of estimating the long run covariance matrix.

The choice of the small sample boostrapped critical values, instead of the asymptotic ones, is crucial for inference in the case of the parametric estimator for the long-run covariance: the small sample critical values are consistently higher than the asymptotic ones. Consequently, the asymptotic critical values would indicate support for PPP only in the post-73 EC12 panel, while small sample critical values give support for PPP across the board. With the non-parametric correction, both sets of critical values give support for PPP for panels with the full 48-year-sample (even at the 10% level). For the shorter post 1973 panel, PPP gets supported only at 1% level of significance with the asymptotic critical values.

Overall, the NH test gives remarkable support for the purchasing power parity. Also, remembering that the test accounts for cross-sectional correlation, and is invariant to the benchmark currency, the results seems even more remarkable.

4.3 Robustness check

Undoubtedly, the support for PPP could be due to the low power of the NH test. To guard against this possibility, we study the power performance of the test in the following manner. Under the null, AR(p) models for p=1 to 4 are fitted and the order for each series is chosen using the BIC as the selection criteria. Under the alternative, ARIMA (p,1,0) models are fitted, and the order for each series is chosen similarly. Taking these AR orders as given, SURE estimation is applied to the panel to get estimates for the individual coefficients and the error covariance matrix both under the I(0) null and the I(1) alternative. The estimated best-fitting regressions for the panel are then

[14] In the full panel with post 1973 data the estimation runs into problems due to the fact that the number of cross sectional units is very close to the number of time periods ($N = 23$; $T = 24$). Therefore, no post Bretton Woods results are available in this case.

used to simulate 5000 time series for each of the 3 panel dimensions. The different methods of estimating the long-run covariance are applied to the generated panels, and the distributions of the test statistics are calculated under the null and the alternative. The farther away the distributions are from each other, the better power we can expect.

The power performance with some other alternatives is also of interest to be investigated, especially those with some of real rates being unit roots and some not. The combinations from the 12- or 23-country panel, however, are too numerous to be exhaustive. Here, we emphasize the information content of the data and give the alternative under consideration good chance of success by letting the data decide on what the best fits are.

Table 2 reports 3 summary numbers that help us to describe the power of the tests to distinguish the alternatives. The distribution of the test statistic under the null is always to the left of the distribution under the unit root alternative. The numbers at `pwr1' column in Table 2 report the probability of rejecting the null at the 5% small sample significance level when the alternative is true. `pwr2' gives the same probability at the sample statistic. The last column (p-val) gives finally the area of the null distribution that is to the left of the sample statistic.

Generally, the power of the test with the non-parametric correction to distinguish alternatives is quite poor. On the other hand, for the parametric correction, the power is very good. The results with parametric correction are thus more reliable in drawing inference on stationarity of deviations from PPP. For any panels considered, the probabilities of rejecting the alternative at 5% level by the test with parametric correction is never less than 81% when the estimated unit root panel processes are true. This indicates that the test with parametric correction is indeed very capable of distinguishing between the two alternatives. Further, we could not reject the estimated stationary panel processes by the test with parametric correction at any conventional significance levels. A level as high as 47%, 23%, and 95% is needed for a rejection, respectively, for the shorter EC12, for the longer EC12, and for the whole panel over 1949-96. These results simply imply that the sample value of the NH test is much more likely to come from the estimated stationary panel processes than from the estimated unit root panel ones.

Are the results with the panels over 1949-96 robust to the change in exchange rate regime?

Whether the real exchange rate is a stationary or a nonstationary process can make quite a difference for long-run forecasting.[15] With longer time series, the difference would be more significant as the parameters of the processes can be estimated more precisely. But the premise is that there is no structural shift in the series. We examine the possibility of a structural shift as follows. The power of the NH-test against the alternative of a regime shift process is studied the same way we studied its power against a unit root process previously. Under the alternative, AR(p) models with an intercept jump at 1973 are fitted to each series and simulated. AR(p) models with other types of structural shifts are also fit to the data, including those with jumps on the whole and partial regression coefficients. These are intended to capture the notion that speeds of adjustment to PPP may vary with the exchange rate regimes. AR(p) with an intercept shift, however, is chosen as the best fit in terms of significance of estimated coefficients.

Table 3 presents the results with power of the test to discriminate the alternative with a change in intercept. As in Table 2, the test with parametric correction continues to exhibit very good power. But it is not the case again for the one with non-parametric correction. The results here appear to preclude the relevance of the exchange rate regimes to the behavior of real exchange rates. There is a clear evidence in favor of a stable stationary process spanning 1949-96 as a true characterization of real exchange rate behavior.

In summary, we can conclude that there is little evidence against PPP hypothesis. We cannot hope to ultimately decide whether any time series processes are I(0), I(1) or regime-dependent in finite samples. This is reflected in the relatively poor power performance of the NH test in some cases when the best fittings with a unit root and with an intercept shift are used as alternatives. However, the results are remarkably consistent with stationary real exchange rates in all the panels and across the different methods of adjusting for cross-sectional dependence and serial correlation when calculating the long-run covariance matrix.

[15] Lothian and Taylor (1996) and Kuo and Mikkola (1999) show that using a stationary specification, RMSE tends to be smaller as the forecasting period lengthens in the univariate context.

5 Comparison to previous results

Is the support for the PPP specific to the NH-testing methodology or the particular data set? Would the conclusions stand with the standard LLC test, and how do the results relate to the previous literature with the unit root null? What about the power of the LLC test to distinguish our best stationary and the unit root panel fittings from each other? These are the questions we focus lastly.

We start by applying the standard testing exercise with the unit root null. Attention is paid to the effects of the treatment of serial correlation, heteroscedasticity and cross-sectional dependence arising due to the common benchmark currency. These are factors that may explain some of the contradictory results found in the previous papers. The following panel, which is practically identical to eq. 1; is estimated

$$\Delta \tilde{q}_{it} = \rho \tilde{q}_{i,t-1} + \sum_{L=1}^{p} \phi_{iL} \Delta \tilde{q}_{it-L} + \varepsilon_{it} \qquad \text{for } i = 1, \ldots, N \tag{6}$$

where $\tilde{q}_{it} = q_{it} - \frac{1}{T}\sum_{t=1}^{T} q_{it}$. This estimation has been done previously by e.g. Oh (1996), Wu (1996) and Lothian (1997), who all make use of the LLC test.

The LLC test depends crucially on the assumption of independence across individuals. In the case of real exchange rates, this can be a potentially serious limitation as noted earlier. Since there are no available unit root tests that account for the economic dependence across the countries, we focus in this section on the relevance of the cross-sectional dependence arising because of a common benchmark currency. To do this, we simply deduct the time specific means by replacing $\tilde{q}_{it}$ by $\tilde{q}_{it} - \frac{1}{N}\sum_{i=1}^{N} \tilde{q}_{it}$ in (6).[16]

5.1 Results with the LLC test

The upper part of Table 4 reports the LLC statistics for our panel of 23 industrial countries and the 12 EC countries, over 1949-96 and post 1973 separately.

Sensitivity of results to the treatment of serial correlation is studied by doing the panel estimations separately for $p = 0, 1, 2$, thus allowing for up to 2 lagged differences in the estimated

[16] This is also done in e.g. Wu (1996) and Coakley and Fuertes (1997).

model.[17] In Table 4, the three rows in each case correspond to $p = 0; 1; 2$. The importance of heteroscedasticity is checked by doing the estimations separately for a model which assumes equal error variance in all countries (het=0 in the Table) and for a model which normalizes according to country specific error variances (het=1). Finally, the models are estimated ignoring the cross sectional dependence due to common benchmark (with only country specific means deducted) and allowing for cross-sectional dependence due to common benchmark (with country and time specific means deducted).

When cross-sectional dependence is not accounted for, for the full panel (1949-96), the unit root can be rejected at the 1 or 5% level when ADF-form is used, regardless of whether heteroscedasticity is accounted for or not. For the EC12 panel over the entire period, the results are in line with those with the full panel. It is noteworthy that when serial correlation is not corrected for the unit root cannot be rejected. For the EC12 over the post-Bretton Woods, the unit root can then be rejected at the 1% level. This is a rejection similar to that found in MacDonald (1996) who has 3 countries different from ours. There is a clear support for PPP for the EC12 countries.

Table 4 gives a puzzling message on the effects of cross-sectional dependence. Once the time specific means are deducted, as discussed above, to account for cross-sectional dependence arising from common benchmark currency, the null of unit root can no longer be rejected for the full sample for $p = 1$ or $p = 2$.[18] However, when serial correlation is not accounted for, the unit root can be rejected.[19] In the smaller panel of the EC12, accounting for cross-sectional dependence has no effect on the conclusions: the null can be rejected at the 1% level. This holds again for both the entire period and the post-Bretton Woods sample. The results seem to pose little evidence against PPP.

[17] In our full panel, using the Akaike (AIC) information criteria or the BIC to choose a best fitting AR(p) model for each individual series gives the following results. BIC (AIC) chooses AR(2) for 14 (18) countries, AR(1) for 8 (3) countries and AR(4) for the remaining 1 (2) countries. Frequently in the literature, an AR(1) or AR(2) model is fitted to the real exchange rates

[18] It requires a 20% level of significance, approximately -6.6, for rejection.

[19] This raises the question on what value for p should be used. E.g. Wu (1996) fits an AR(2) model to the data also when time specific means are deducted. In this case, however, it is reasonable to think that the serial correlation properties may play a role here. Indeed, frequently a lower order AR fits the data better with time specific means subtracted. BIC (AIC) now chooses AR(2) for 7 (10) countries, AR(1) for 13 (7) countries and AR(3) or AR(4) for the remaining countries. Compared to the original data the two criteria disagree more frequently, and the chosen p's vary more across countries. It is no longer clear that the AR(2) model is the one to be chosen automatically.

5.2 Accounting for different results in the previous literature

Since there are numerous studies using similar sets of countries and methods, yet arriving at different results, it is useful to gauge the consequences of the differences in their treatment of serial correlation, heteroscedasticity and cross-sectional dependence. It indeed turns out that some of the seemingly contradictory results are explained by different treatment of these issues in the panel setup.

We do separate estimations where our dataset is split to conform as closely as possible, in terms of both sample period and selection of countries to the following studies: MacDonald (1996), Oh (1996) and Wu (1996). These studies are chosen as an example, since they all use annual data and somewhat similar set of countries as we do, thus allowing for comparison. At the same time, they differ in their treatment of serial correlation, heteroscedasticity and cross-sectional dependence.[20]

The bold numbers in the table refer to the estimates that are most comparable to the ones reported in the respective studies in terms of their treatment of serial correlation, heteroscedasticity and cross-sectional dependence. For the Oh sample (1960-89), the results are strikingly similar to Oh's, even though he uses Penn data.[21] Oh's estimates for ρ were -17 (t-value=-10.74) and -18 (t-value=-9.79) when p=1 and 2 respectively. Thus, our estimation agrees with Oh's in rejection the null of unit root at the 1% level. Accounting for cross-sectional dependence does not reverse his support for PPP. The support for PPP found in MacDonald does not disappear after benchmarking is controlled for. Likewise, Wu's result of rejecting the null with time specific means deducted is found to stand with the estimated AR models, different from his AR(2) ones. These estimations seem to give additional support to PPP by confirming that the results of Oh, Wu and McDonald's are not sensitive to their treatment of serial correlation, heteroscedasticity or cross-sectional dependence due to benchmarking.

Overall, the LLC test statistics are somewhat smaller in absolute terms when cross-sectional dependence arising from benchmarking is accounted for. Drawing conclusions to the extent of PPP

[20] To treat serial correlation Wu fits an AR(2) model to the errors, Oh estimates separate regressions for lags from 1 to 4 and MacDonald allows for differing lag lengths for different countries. MacDonald corrects for heteroscedasticity, while Oh and Wu do not. MacDonald and Oh do not account for cross-sectional dependence, due to common benchmark currency while Wu does.

[21] Here, we have exactly the same set of countries and the sample period.

not holding is, however, not warranted. It should also be noted that even if the asymptotic critical values are not affected by the estimation of aggregate common factors, it is possible that the finite sample properties are dependent on this panel testing procedure. We will next simulate the finite sample critical values, given our best-fitting panel regressions.

5.3 Small sample properties of LLC test

We turn next to small sample simulations to look at the capability of the LLC test to distinguish between the alternative best stationary and non-stationary panel fittings. We focus simulations on the EC 12 panel over both 1949-96 and post-1973. Panel A of Table 5 reports the small sample power of the test, and Panel B reports the small sample critical values for the panels. The experiment is identical to the one done with the NH test previously.

Applying the small sample critical values does not affect the previous basic conclusion: the null of unit root can be rejected for the EC 12 panel over 1949-96 and 1973-96. The null can be always rejected at the 5% level of significance. Some summary information on the location of the distributions of the simulated test statistics under the null and the alternative are reported in the table. They indicate that the LLC test statistic in our sample of EC countries is indeed more likely to come from the stationary distribution. It is worth noting that correcting for the benchmarking effect reduces the overlap of the distributions of the test statistic under the alternatives. This indicates the increased power of the test. In fact, in the presence of cross-sectional dependence, accounting for it by subtracting the time specific means yields the distributions with mass more concentrated toward means and thinner tails. Practically, this shifts the critical values rightwards, and decreases the p-value as evident from the table. Failing to control for cross-sectional correlation, as a result, may hamper the power of the test, and lead to a spurious acceptance of PPP as compellingly argued by O'Connell (1998).

6 Concluding Comments

The purchasing power parity hypothesis cannot be rejected for our panel of 23 industrial countries, nor for the subsample of 12 European community countries over 1949-96. PPP is found to hold for the European countries on the post Bretton Woods period as well. The application of the Nyblom-

Harvey (1998) test allowed us to test the null hypothesis directly. Additional complementary evidence in favor of PPP comes from applying the panel version of conventional unit root test. The panel approach adopted here, thus, points toward the real exchange rates of these countries as mean-reverting, despite of a great deal of short-term variations.

The findings, moreover, stands with cross-sectional correlation and regime changes as the test speaks directly to the issues. The extreme finding of O'Connell (1998) that there is no support for PPP when the cross-sectional dependence is accounted for, therefore, is not present in our study. In contrast to most recent panel studies, the evidence is based on cross-country data spanning both the current float and the previous regime. It reinforces the view that PPP, as a long run phenomenon, has little to do with the exchange rate regimes. Indeed, our evidence is not any weaker for the longer panel of European countries than for the shorter counterpart.

The findings with longer panels here, therefore, lie between the recent panel evidence under the float and what have been learned from the studies using long time series. The behavior of real exchange rates, in view of this body of evidence, appear to be stable and similar, at least concerning the stability of the first moments. While empirical validity of this view is yet to be investigated, the evidence presented in the paper, as in the recent literature, reveals that purchasing power parity remains to serve us well as a first long run approximation. The PPP hypothesis, as it seems, after all finds its voice echoing.

References

[1] Abuaf, N. and P. Jorion, 1990, Purchasing Power parity in the Long Run, Journal of Finance 45, 157-174.

[2] Canova, F. and H. Dellas, 1993, Trade interdependence and the international business cycle, Journal of International Economics 34, 23-47.

[3] Choi, I. and B.C. Ahn, 1999, Testing the null of stationarity for multiple time series, Journal of Econometrics 88, 41-78.

[4] Coakley, J. and A.M. Fuertes, 1997, New panel unit root tests of PPP, Economic Letters 57, 17-22.

[5] Den Haan, W. and A. Levin, 1997, A Practitioner's Guide to Robust Covariance Matrix Estimation, forthcoming in Handbook of Statistics.

[6] Frankel, J. and A. Rose, 1996, A Panel Project on Purchasing Power Parity: Mean Reversion within and between Countries, Journal of International Economics 40, 209-224.

[7] Froot, K., M. Kim and K. Rogoff, 1995, The law of one price over 700 years, National Bureau of Economic Research, Working Paper 5132.

[8] Froot, K. and K. Rogoff, 1995, Perspectives of PPP and Long-Run Real Exchange Rates, in: G. Grossman and K. Rogoff, eds., Handbook of International Economics, vol III (North-Holland, Amsterdam).

[9] Hakkio, G., 1984, A Re-examination of Purchasing Power Parity, A multi-country and multi-period study, Journal of International Economics 17, 265-277.

[10] Im, K.S., H. Pesaran and Y. Shin, 1995, Testing for Unit Roots in Heterogeneous Panels, University of Cambridge, Department of Applied Economics Working Paper No. 9526.

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

[12] Koedijk, K., P. Schotman and M., Van Dijk, 1998, The re-emergence of PPP in the 1990s, Journal of International Money and Finance 17, 51-61.

[13] Kuo, B.-S., 1998, Test for partial parameter instability in regressions with I(1) processes, Journal of Econometrics 86, 337-368.

[14] Kuo, B.-S. and A. Mikkola, 1999, Re-examining long-run purchasing power parity, Journal of International Money and Finance, forthcoming.

[15] Kwiatkowski, D., P. Phillips, P. Schmidt and Y. Shin, 1992, Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root?, Journal of Econometrics 54, 159-178.

[16] Levin, A., C.-F. Lin and C.-S. Chu, 1997, Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties, photocopy, earlier version by Levin and Lin as a UCSD Working Paper, 93-56.

[17] Leybourne, S J. and B P M. McCabe, 1994, A Consistent Test for a Unit Root, Journal of Business & Economic Statistics 12, 157-166.

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

[19] Lothian, J., 1997, Multi-country evidence on the behavior of purchasing power parity under the current float, Journal of International Money and Finance 16, 19-35.

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

[21] MacDonald, R., 1996, Panel unit root tests and real exchange rates, Economic Letters 50, 7-11.

[22] Nyblom, J. and A. Harvey, 1998, Testing for Common Stochastic Trends, Econometric Theory, forthcoming.

[23] O'Connell, P., 1998, The Overvaluation of Purchasing Power Parity, Journal of International Economics 44, 1-19.

[24] Oh, K.-Y., 1996, Purchasing Power Parity and unit root tests using panel data, Journal of International Money and Finance 15, 405-418.

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

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

[27] Rogoff, R., 1996, The purchasing power parity puzzle, Journal of Economic Literature 34, 647-668.

[28] Shiller, R. and P. Perron, 1985, Testing the Random Walk Hypothesis: Power versus Frequency of Observation, Economics Letters 18, 381-386.

[29] Stockman, A., 1987, The Equilibrium Approach to Exchange Rates, Federal Reserve Bank of Richmond Economic Review 73, 12-31.

[30] Stockman, A. and L. Tesar, 1995, Tastes and Technology in a Two-Country Model of the Business Cycle: Explaining International Comovements, American Economic Review 85, 168-185

[31] Taylor, M. and L. Sarno, 1997, The Behaviour of Real Exchange Rates during the Post-Bretton Woods Period, Centre for Economic Policy Research, Discussion Paper No. 1730.

[32] Wei, S. and D. Parsley, 1995, Purchasing power dis-parity during the floating rate period: Exchange rate volatility, trade barriers and other culprits, NBER Working Paper no. 5032.

[33] Wu, Y., 1996, Are Real Exchange Rates Nonstationary? Evidence from a Panel-Data Test, Journal of Money, Credit and Banking 28, 54-63.

Table 1

Panel A : NH test statistics

LR cov. estimator	EC12 (N = 12) 1949 ¡ 96	EC12 (N = 12) 1973 ¡ 96	FULL PANEL (N = 23) 1949 ¡ 96
Non-parametric			
m = 4	1 90 ***	2 24 ***	2 91 ***
m = 8	2 11 **	3 22 ***	3 98 ***
m = 12	2 27 ***	4 02 ***	4 .76 ***
Parametric	4 18 ***	2 58 ***	5 84 ***
No correction	7 07	3 66 ***	7 59 ***

Note:

1. Nonparametric estimates for the long run covariance is obtained using a Bartlett kernel. m represents the bandwidth number.
2. ***; ** and * denote the cases where the null of PPP cannot be rejected at the 10, 5 and 1% level of significance using the small sample critical values in Panel B.

Panel B : Small sample and asymptotic critical values

LR cov. estimator	N = 12, T = 48			N = 12, T = 24			N = 23, T = 48		
Nonparametric	10	05	01	10	05	01	10	05	01
m = 4	1 92	1 93	1 95	2 28	2 34	2 4.4	2 96	3 00	3 10
m = 8	2 09	2 13	2 21	3 22	3 26	3 34	4 11	4 16	4 24
m = 12	2 62	2 65	2 .73	4 05	4 08	4 16	5 03	5 07	5 15
Parametric	4 95	5 35	5 92	3 15	3 25	3 35	6 58	6 67	6 .78
No correction	6 91	7 00	7 05	3 67	3 69	3 .71	7 .70	7 .70	7 .71
Asymptotics	2 69	2 94	3 51	2 69	2 94	3 51	4 81	5 15	5 .78

Note:

1. The finite-sample critical values are simulated as follows. The best AR order for each series in the panel is chosen by fitting AR(p) processes for p = 1 to 4, and using the BIC as the selection criteria. Taking these AR orders as given, SURE estimation is applied to the panel to estimate the individual coefficients and the error covariance matrix. The estimated processes for the panel are then used to simulate 5000 time series for each the 3 panel dimensions. The different methods of estimating the long-run covariance are applied to the generated series, and the critical values associated with the top 10, 5 and 1% of the values of the test-statistics are calculated.
2. Asymptotic critical values are calculated from an approximation to the limiting distributions, CVM (12) and CVM (23) in (3) by truncating the summation at k = 5000, and simulating 8000 points from the distribution.

Table 2 Power of the NH-test against the best-fitting unit root process

LR cov. est.	EC12 (N = 12) 1949-96			EC12 (N = 12) 1973-96			FULL PANEL (N = 23) 1949-96		
Nonparametric	pwr1	pwr2	p-val	pwr1	pwr2	p-val	pwr1	pwr2	p-val
m = 4	15	58	.78	05	17	83	07	26	.79
m = 8	07	10	93	05	10	90	05	41	59
m = 12	08	97	05	06	16	86	05	83	18
Parametric	99	100	.77	81	100	53	99	100	05
No correction	99	95	99	80	93	86	02	100	00

Notes:

1. 'pwr1' and 'pwr2' denote, respectively, the power of the test statistic at the 5% finite sample significance level and at the sample statistic.

2. 'p-val' denotes the percentage of the distribution of the test statistic (under the null) that is to the left of the sample statistic.

Table 3 Power of the NH-test against the best-fitting process with an intercept jump

LR cov. est.	EC12 (N = 12) 1949-96			FULL PANEL (N = 23) 1949-96		
Nonparametric	pwr1	pwr2	p-val	pwr1	pwr2	p-val
m = 4	13	69	53	07	84	20
m = 8	07	04	97	05	96	04
m = 12	06	98	02	05	100	00
Parametric	64	85	83	83	99	57
No correction	01	00	100	00	00	100

Notes: See Table 2.

Table 4 LLC test statistics and comparison to previous results

The Estimated Model: $\Delta \tilde{q}_{it} = \rho \tilde{q}_{i,t-1} + \sum_{L=1}^{p} \phi_{iL} \Delta \tilde{q}_{i,t-L} + \varepsilon_{it}$

		Country specific means				Time specific means			
		het=1		het=0		het=1		het=0	
	N	$\hat{\rho}$	t-val.	$\hat{\rho}$	t-val.	$\hat{\rho}$	t-val.	$\hat{\rho}$	t-val.
1949-96	23	-08	-6.7	-09	-72*	-07	-73**	-10	-92***
Full Panel		-11	-93***	-10	-8.7***	-07	-6.7	-07	-65
		-09	-80**	-09	-75**	-06	-61	-06	-60
1949-96	12	-10	-58**	-12	-6.7***	-13	-7.7***	-24	-13.7***
EC12		-11	-68***	-11	-68***	-10	-68***	-12	-75***
		-11	-64***	-11	-64***	-10	-64***	-13	-7.7***
1973-96	12	-21	-55*	-20	-55*	-18	-54*	-17	-52
EC12		-29	-78***	-30	-79***	-21	-69***	-22	-70***
		-32	-7.7***	-32	-7.7***	-23	-71***	-23	-69***
1973-96	23	-21	-78**	-21	-78**	-20	-78**	-20	-78**
(MacDonald)		-30	-115***	-30	-111***	-25	-98***	-26	-98***
		-32	-106***	-32	-105***	-25	-8.7***	-26	-88***
1974-92	18	-19	-55	-19	-55	-23	-6.7*	-24	-68*
(Wu)		-29	-90***	-30	-89***	-30	-88***	-31	-8.7***
		-36	-95***	-35	-88***	-30	-74***	-34	-84***
1960-89	23	-13	-73**	-13	-72*	-10	-63	-09	-60
(Oh)		-20	-120***	-19	-113***	-12	-80**	-11	-73**
		-19	-100***	-17	-94***	-10	-63	-10	-59

Note:

1. The three numbers reported in a column under each sample refer to estimations with different number of lagged differences included. The first number corresponds to no lagged differences, the second to $p = 1$ and the third to $p = 2$.
2. For the panels considered here, the significant rejections of the unit root are indicated by *, ** and *** at the 10, 5 and 1% level respectively. For N = 23, the asymptotic critical values corresponding to the 10, 5 and 1% level are -70, -73 and -80, while for N = 12, they are -54, -5.7 and -63. Wu (1996) reports the critical values simulated to correspond to his panel dimensions (-6.7, -70 and -76), which are used here.
3. het= 1 (het= 0) indicates estimating the model with (without) correction for heteroscedasticity.
4. Country specific means corresponds to $\tilde{q}_{it} = q_{it} - \frac{1}{T}\sum_{t=1} q_{it}$ and time specific means further subtracts the time specific means from the data to arrive at $\tilde{q}_{it}$:
5. Bold numbers indicate those that correspond most closely to the estimations in the papers mentioned.
6. The Wu sample drops the following countries to conform with the Wu choice of countries: Ireland, Malta, Australia, New Zealand and South Africa.

Table 5

Panel A: Power of the LLC test against the best fitting stationary process

Correction	EC12 (N = 12) 1949-96 pwr1	pwr2	p-val	1973-96 pwr1	pwr2	p-val
None	.49	.69	.11	.21	.79	.32
+ serial correlation.	.47	.78	.21	.24	.47	.14
+ heteroscedasticity	.46	.78	.22	.23	.47	.16
+ cross-sectional dep.	.78	.77	.05	.61	.58	.04

Notes:

1. 'pwr1' and 'pwr2' denote the power of the test-statistic at the 5% finite sample significance level and at the sample statistic. The sample statistic calculated with all the correction is -6.45 for EC12 over 1949-96, and -7.08 for EC12 over 1973-96.

2. 'p-val' denotes the percentage of the distribution of the test statistic (under the null) that is to the left of the sample statistic.

Panel B: Small sample critical values of the LLC-test

Correction	EC12 (N = 12) T = 48 .10	.05	.01	T = 24 .10	.05	.01
None	-6.74	-7.56	-9.21	-7.44	-8.38	-10.4
+ serial corr.	-7.25	-7.94	-9.37	-8.19	-9.08	-11.0
+ het.	-7.35	-8.01	-9.48	-8.29	-9.20	-11.2
+ c-s. dep.	-6.01	-6.40	-7.28	-6.50	-6.95	-7.95

Note: For the method of simulating the critical values see notes to Panel B of Table 1.