How sure are we about purchasing power parity? Panel evidence with the null ...
Biing-Shen Kuo; Anne Mikkola
Journal of Money, Credit, and Banking; Aug 2001; 33, 3; ABI/INFORM Global
pg. 767

BIING-SHEN KUO
ANNE MIKKOLA

How Sure Are We about Purchasing Power Parity?
Panel Evidence with the Null of Stationary Real Exchange Rates

This article presents evidence on mean reversion in industrial countries' real exchange rates in a setup that accounts naturally for cross-sectional dependence, is invariant to the benchmark currency, and actually tests for the null of interest, that is, purchasing power parity. Our results are based on the Kwiatkowski et al. (1992) test for the stationarity null generalized in a multivariate random walk plus noise model by Nyblom and Harvey (2000).

For purchasing power parity (PPP) to hold in the long run, real exchange rates must be stationary. Permanent shocks to them would imply a permanent tendency for the purchasing power of the currencies to deviate from one another. Whether real exchange rates are stationary or nonstationary 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 long-run behavior. This is the view of practicing 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 real exchange rates would be problematic for many of the established theories. Furthermore, it would make long-run forecasting a useless exercise.

Having said this, does this mean we believe in the possibility of an “eternal” equilibrium real exchange rate? Few economists would go that far. There is evidence that price levels in rich countries tend to be higher than in poor countries when converted to a common currency, due to, for example, the Balassa-Samuelson effect. The evi-

The authors wish to thank two anonymous referees and the editor Nelson Mark for their helpful comments and suggestions, which enhanced the readability of the paper. They also gratefully acknowledge grants from the Academy of Finland and the Yrjö Jahnsson Foundation to the Research Unit on Economic Structures and Growth (RUESG) and from the National Science Council (NSC-88-2415-H-004-010). This paper was finalized while the second author was visiting the Bank of Finland.

Biing-Shen Kuo is associate professor in the Graduate Institute of International Trade, National Chengchi University. E-mail: bskuo@nccu.edu.tw Anne Mikkola is a postdoctoral fellow in economics at the University of Helsinki. E-mail: anne.mikkola@helsinki.fi

Journal of Money, Credit, and Banking, Vol. 33, No. 3 (August 2001) Copyright 2001 by The Ohio State University

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
dence for the industrial countries is more debatable, however (see Rogoff 1996). Thus, what we are really testing empirically is whether 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 real exchange rates.

Since PPP holds at best only as a long-run relationship, statistical inference on it depends critically on the number of available observations. Therefore, we consider it very important to use as much data as possible. Many of the PPP studies done in the past were based on a single time series due to the lack of suitable panel methods for testing. Tests for unit roots in panels have been developed only in the 1990s (see Im, Pesaran, and Shin 1997; Levin, Lin, and Chu 1997, hereafter LLC1). For the PPP tests, this has meant testing for the “wrong” null, that is, the null of the theory not holding. Empirical testing on relatively short univariate time series, for example, on post-1973 data, typically failed to support PPP. The emerging “consensus” of the failure of PPP started to shift back toward its acceptance in the 1990s. This occurred when studies using longer time series or panel methods were able to reject the unit root. Somewhat remarkably, both approaches arrived at similar speeds of adjustment to PPP, the half-life of deviations from PPP being in the range of three to five years [for discussion on the empirical results other than the very latest ones, see Froot and Rogoff (1995) for an excellent survey; for panel studies, see, for example, Coakley and Fuertes (1997), MacDonald (1996), Oh (1996), Wei and Parsley (1995), and Wu (1996)]. Support for PPP was considered to be due to increased power of the unit root tests with more observations and more variation in data.

However, one should be very careful in drawing conclusions based on these tests. The inference can be misleading since little is known about their capabilities of distinguishing the relevant alternatives in particular real-life time series. In testing for PPP, caution is needed in statistical decision making because the two competing processes are very similar in any finite samples, one being highly persistent but stationary and the other nonstationary.

Some recent studies indicate that the problems associated with the standard testing procedures are greater than is perhaps recognized. It is known that unit root tests have low power against persistent alternatives in small samples. In addition, there are serious size and power problems that can lead to faulty inference in either direction. Engel (2000) shows that while both the unit root and stationarity tests indicate simultaneous support for PPP (for the hundred-year U.K./U.S. real exchange rate series), the truth may be the opposite. His simulations suggest the possibility of simultaneous faulty inference with both tests, if the real exchange rate consists of a stationary and a random walk component. In this case, the unit root tests over-reject, while the KPSS test (Kwiatkowski, Phillips, Schmidt, and Shin 1992) for stationarity has extremely low power to detect the random walk component. Thus, if there is a random walk component, it may not be detected by any of the tests. Rather the re-

1. Earlier version of the paper was by Levin and Lin (1993). The revision is concerned with proofs of the asymptotic results. It does not affect performance of the test or critical values derived in the earlier version.
results with the alternative nulls may appear to give reinforcing evidence for PPP. Simulation results by Caner and Kilian (1999) point in the opposite direction: if the true process is stationary but persistent, the KPSS test for stationarity rejects the null too frequently. Increasing the sample size is of little help.

Inference in the panel setup involves its own considerations and problems with regard to interpretation. Power of the panel unit root tests is generally considerably higher than that of the univariate counterpart. Im, Pesaran, and Shin (1997), however, show that power can drop quite dramatically in the LLC test if the lag order is overfit in performing the test. On the other hand, Monte Carlo simulations by Taylor and Sarno (1998) illustrate, in a revealing manner, how uninformative the rejection of the null of panel unit root tests can be about PPP: the null hypothesis that all the series are realizations of unit root processes is violated even if only one of the series is stationary. The simulations show that 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 the null is rejected. To facilitate interpretation, one would prefer to have the stationarity as the null to be tested. This is the approach applied in this paper.

The aim of this paper is to judge the extent to which we can think of PPP holding, given the problems in statistical decision making that are inherent in the tests. Since the power and size distortion problems make judging the reliability of inference all but impossible in practice, we suggest an alternative, or rather complementary, route for testing. The idea is to utilize to a larger extent the information content of the data in distinguishing the relevant alternatives. We fit alternative stationary and nonstationary processes to the data and apply finite-sample simulation techniques to draw conclusions on which one of them the test statistic is more likely to come from. This gives a clearer picture of the capability of the tests to tell the alternatives apart. The reliability of our approach is limited by how well the two alternative fitted processes describe the actual alternatives. In empirical studies, a simple AR(p) process is typically found to describe real exchange rates well. We take this as our specification when PPP holds and real exchange rates are stationary. With PPP not holding, we fit an AR process to the differenced series, thus implying a unit root. Then we generate artificial time series that follow these estimated processes, and simulate small-sample distributions of the test statistics under these two processes. Finally, we calculate the test statistic from the real data, and see which of the two distributions it is more likely to come from. An application of this methodology can also be found in Kuo and Mikkola (1999) for the case of a U.S./U.K. real exchange rate series, and in Rudebusch (1993) for the case of the U.S. GNP series.

We do not want to take the “pick one of the two specifications” approach too far for two related reasons. Firstly, the potential nonstationarity appearing in reality might be a more complex process, which is not well approximated by the fitted ARIMA process even for long-run purposes. Secondly, given that our hypothesis is that PPP holds, we want to be particularly careful that we do not reject the theory too lightly. We see the relevance of our results as giving insights complementary to conducting regular inference. Carrying out regular inference, that is, testing if the null of PPP can be rejected,
is identical to looking at just one of our alternative distributions. In this sense, our exercise only broadens our view of what may be going on behind the scenes.

An ideal test for our purposes has been recently developed by Nyblom and Harvey (2000, hereafter NH). Unlike the standard panel version of the ADF test (the LLC test), the NH test is a test for the null of stationarity. Since PPP is our hypothesis, we really ought to have the stationarity of real exchange rates as the null. It seems unwarranted to draw swift conclusions on PPP not holding on the basis of not being able to reject the unit root—the null that is not our hypothesis. This test has additional advantages that are particularly important in testing for PPP. It incorporates cross-sectional dependence very naturally in the panel estimation. Also, it is invariant to the choice of benchmark currency. This is a desirable property, given that empirical studies tend to give more support for PPP when the German mark, rather than the U.S. dollar, is used as a benchmark currency. We also compare our results to those in the previous literature by repeating our exercise with the LLC test, that is, with the null of unit root.

The test will be applied to a panel of industrial countries’ real exchange rates over the period 1949–1996. Industrial countries are a category used in many previous panel studies, which allows for comparison of the results. It has been argued that the use of data from the pre- and post-Bretton Woods periods is not appropriate in PPP studies because of regime changes. Consequently, much of the empirical work has looked only at data from the floating rate regime after 1973. We, however, think that, as a long-run phenomenon, PPP should apply regardless of the exchange rate regime. Some empirical support for this view is provided in Lothian and Taylor (1996). They show that the stationary process estimated for pre-1973 data performs well in out-of-sample forecasting for the post-73 period as well. Froot, Kim, and Rogoff (1995) 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 seven hundred years in England and Holland. Our results also lend support to this view.

The paper is organized as follows. The choice of data and the estimation of the best-fitting processes are discussed in section 1. The Nyblom-Harvey test and its application to our problem are then discussed in section 2. Results are presented in section 3. Robustness of results to the method of estimation and the choice of the LLC test are discussed in section 4. Section 5 concludes.

1. BEST-FITTING PROCESSES

1.1 Data

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. The real exchange rate for country $i$ at time $t$ is thus

2. It is a multivariate extension of the much-used KPSS test.
\[ q_{it} = p_{it} - e_{it} - p_{USi}, \]

where \( p_{it} \) is the CPI for country \( i \) at time \( t \), \( p_{USi} \) is the CPI for the United States, and \( e_{it} \) is the exchange rate of country \( i \) at time \( t \) in units of country \( i \)'s currency per U.S. dollar.\(^3\) Data is available for the following twenty-four countries over the period 1949–1996: United States, United Kingdom, 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. Two subpanels are used in the estimation: the twelve European countries over the entire sample period and over the post-1973 period, respectively.\(^4\)

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). In annual data, the persistence in real exchange rate movements, as measured by the sum of AR coefficients, is considerably less than in monthly or quarterly data. This makes the statistical distinction between the nonstationary and stationary but persistent alternatives easier: the KPSS test is shown to over-reject the stationarity null, regardless of sample size, when the series is persistent to the degree found in monthly or quarterly real exchange rates (see Caner and Kilian 1999). Since the NH test is a multivariate extension of the KPSS test, this result is likely to hold in our case, thus motivating the use of annual data as well.

1.2 Best-Fitting Stationary and Nonstationary Processes

We start by estimating the best-fitting stationary and difference stationary processes, and their associated error covariance matrices for our panels. These will be the two alternative “true” processes that we attempt to distinguish from each other. In the rest of the paper, we not only apply the NH test to the actual data, but also focus on how well the person applying the NH test would be able to pick the right one of the two alternatives that we estimate in this section.

Only ARIMA processes are considered. For annual real exchange rates, even an AR(1) process is found to work well in describing the stationary alternative (see, for example, Lothian and Taylor 1996). For the nonstationary specification, ARIMA(1,1,0) processes are fitted. They allow for reasonable flexibility in incorporating transitory movements even in the nonstationary specification. Our ultimate goal is to evaluate which of these processes is more likely to describe the data. The appropriate panel tests are applied for that purpose.

To estimate the best-fitting panel processes, we first choose the lag orders for each country and then estimate the full panel given these country-specific specifications.

---

3. 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 65) as the exchange rate. All variables are in logarithms.

4. The EC12 panel includes the United Kingdom, Belgium, Denmark, France, Germany, Italy, Luxembourg, the Netherlands, Greece, Ireland, Portugal, and Spain. Belgium and Luxembourg, though forming a currency union, have different price levels.
The stationary process for country $i$'s real exchange rate is taken to be of the AR($p_i$) form

$$q_{it} = \gamma_i + \sum_{j=1}^{p_i} \phi_{ij} q_{i,t-j} + \nu_{it}.$$  \hspace{1cm} (1)

For the difference stationary specification, an ARIMA($r_i$,1,0) process is fitted to the data,

$$\Delta q_{it} = \sum_{j=1}^{r_i} \lambda_{ij} \Delta q_{i,t-j} + \nu_{i,t}.$$  \hspace{1cm} (2)

For each real exchange rate series, equation (1) and equation (2) are estimated for $p_i$ and $r_i$ varying from one to four. The BIC criteria are used to select the best-fitting specifications, that is, the lag lengths $p_i$ and $r_i$. For the stationary processes, the most common lag length is $p = 2$, which is chosen for 16 and 14 countries in the post-73 and full samples, respectively. One lag is sufficient for 5 and 7 countries, respectively. With the difference stationary specification, the choice of $r_i$ is even more unanimous: $r = 1$ is chosen for 20 and 21 countries, respectively. AIC criteria would choose more lags for only 1 and 2 countries, respectively.

The best-fitting stationary panel is then estimated by taking these $p_i$s as given and applying SURE to the panel where the individual AR($p_i$) processes of equation (1) are stacked. This gives the final panel estimates for $\hat{\gamma}_i$, $\hat{\phi}_{ij}$, and the error covariance matrix. The nonstationary panel is estimated analogously by applying SURE to the stacked relationships of the form of equation (2). To verify the significance of cross-sectional dependence, the likelihood ratio (LR) test is employed to test if the off-diagonal elements in the innovation covariance matrix can be restricted to zero. The values of the $\chi^2$-statistics show clear evidence against no cross-sectional dependence in all the panels under study. We want to account for the cross-sectional dependence in the estimation as far as possible.

An alternative approach to modeling the apparent cross-sectional dependence is to formulate the panels as VAR relationships, where shocks to one country affect the other countries' real exchange rates with a lag. Given that we have large panels with only twenty-four or forty-eight annual observations available, this does not seem feasible here. Also, we might expect the NH test itself to be able to account for the remaining cross-sectional dependence.

---

5. The test is computed as $T[\ln(\hat{\Omega}(0)) - \ln(\hat{\Omega})]$. $\hat{\Omega}(0)$ and $\hat{\Omega}$ 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 for the panel of EC12 countries and a $\chi^2(253)$ distribution for the full dataset of twenty-three real exchange rates. The value of the test-statistic is 1,800 for the full dataset. In the EC dataset for the full sample period and post-1973 period, the values are 1,010 and 697 respectively.
We rule out drift in the unit root, which is consistent with the exclusion of trend in the stationary specification. The theory of long-run PPP is not compatible with trends in the real exchange rates.

2. TESTING FOR PPP: THE NULL OF STATIONARITY

Having stationarity of real exchange rates as the null, rather than the alternative, is desirable for several reasons. It ensures that the null of PPP is not rejected as long as there is no strong evidence against it, thus being consistent with the way classical hypothesis testing is performed. With the standard unit root null in the panel, it is not clear what the rejection of the null implies in terms of support for PPP.

The test developed by Nyblom and Harvey (2000) makes it possible to test for the stationarity directly in the panel setup. This test is too recent to have been applied in empirical work to our knowledge. As a direct multivariate extension of the much used KPSS test, its intuition is familiar, and its performance in the univariate context gives some idea of its performance in the panel context. Performing extensive simulations on its performance in the panel setup becomes an overwhelming task. However, our data-specific simulations do not indicate any unexpected problems with it. For testing PPP, this test seems ideal. We will discuss first the idea and application of the test in our panels as well as its particular advantages in the PPP literature.

2.1 Nyblom-Harvey Test

For illustration, we first present the model in its simplest form with no serial correlation. The panel of real exchange rate series forms a multivariate model of random walk plus noise,

\[ q_t = \delta + \mu_t + \epsilon_t, \quad \epsilon_t \sim iid(0, \Sigma_{\epsilon}) \]

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

where

\[ q_t = (q_{1t}, q_{2t}, \ldots, q_{Nt})', \delta = (\delta_1, \delta_2, \ldots, \delta_N)', \text{ and } \mu_t = (\mu_{1t}, \mu_{2t}, \ldots, \mu_{Nt})'. \]

where \( q_t \) is the vector containing \( N \) real exchange rate time series, \( \delta \) is the nonzero intercept vector whose elements vary with countries but not with time, and \( \mu_t \) is a random-walk time series. The test statistic is

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

where \( S \) is an estimate for the so-called long-run error covariance matrix computed as
\[ S = T^{-1} \sum_{t=1}^{T} (q_t - \overline{q})(q_t - \overline{q})', \quad \text{where} \quad \overline{q} = \frac{1}{T} \sum_{t=1}^{T} q_t, \]

and

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

As in the KPSS test, \( \xi_N \) is a test for the null of stationarity. Stationarity will be rejected for large values of \( \xi_N \). 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 the real exchange rates are stationary. Under the alternative, all the real exchange rates, or part of them, have unit roots. This is exactly the desirable setup for making inferences on PPP.

The limiting distribution of \( \xi_N \) under the null is

\[ \xi_N \to^d \int_0^1 B(u)'B(u)du = \sum_{k=1}^{\infty} (\pi k)^{-2} u_k'u_k, \quad (4) \]

where \( u_k \sim NID(0, \Sigma_\eta) \) 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, \( N \).

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 individual squared partial sums of each series (the diagonal elements in \( C, \left[ \sum_{t=1}^{j} (q_{kt} - \overline{q}_k) \right]^2 \)), 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 \( C, \left[ \sum_{t=1}^{j} (q_{kt} - \overline{q}_k) \right] \left[ \sum_{l=1}^{j} (q_{lt} - \overline{q}_l) \right] \)). This enables the NH test to have more opportunities to detect the stationarity property in the data than the KPSS test.

6. Nyblom and Harvey tabulate the critical values only for up to \( N = 4 \). We compute the critical values for \( N \) greater than 4 by following the formula in (4) with a sufficiently large truncation number to approximate the distribution well enough.

7. Choi and Ahn (1999) mount tests for the stationarity null for multiple time series. One of them, de-
2.2 Accounting for Serial Correlation

For the test to be useful in practice, we need to allow for serial correlation. Because the NH test is constructed under the assumption of error independence, the estimation now has to account for autocorrelation. This is done by applying a parametric correction to our individual series in the panel.\(^8\) Then the NH test is applied to the panel of the real exchange rates data from which serial correlation is first being removed.

The model for an individual country \(i\)'s real exchange rate at time \(t\) allowing for autocorrelation is now written as

\[
\alpha_i(L)q_{it} = \delta_i + \mu_i + \epsilon_{it},
\]

where \(\alpha_i(L) = 1 - \alpha_{i1}L - \ldots - \alpha_{ip}L^p\) is an 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. The model also allows for various convergence speeds across countries as no restriction is placed on the AR coefficients.

For the test to have power, the AR parameters need to be estimated in such a way that they are consistent under both the null and the alternative. To achieve this, we conduct the estimation strategy on the reduced form of model (5) as suggested by Leybourne and McCabe (1994). The reduced form is obtained by differencing the structural model in (5):

\[
\alpha_i(L)\Delta q_{it} = (1 - \theta_iL)\zeta_{it},
\]

where \(\alpha_i(L)\) is as defined before, and \(\sigma_{\zeta}^2\) is distributed as iid\((0,\sigma_{\zeta}^2)\).\(^9\) This reduced form allows estimation of \(\alpha_i(L)\) consistently under both the null and the alternative.

Testing for stationarity is now equivalent to testing if \(\theta_i = 1\). Thus, under \(H_0: \theta_i = 1\), \(q_{it}\) in equation (6) is a stationary AR\(p_i\) process, and under \(H_a: |\theta_i| < 1\), it is a non-stationary ARIMA\((p_i,1,1)\) process.

We proceed to fit an ARIMA\((p_i,1,1)\) process to \(q_{it}\) that is, an ARMA\((p_i,1)\) to \(\Delta q_{it}\), as in equation (6). The AR orders for each country are taken from the best-fittings as described in section 1. Strictly speaking, we should not take the AR orders as given, since the person using the NH test does not know the true order. However, for the annual data, the lag orders seem to be quite similar across countries and vary little across different criteria (as reported in section 1). Therefore, we decided to use the true orders, rather than estimate the orders separately, for each series in all the thousands of simulations that we will be doing. To make sure this simplification is not

---

8. In the univariate context, the parametric correction appears to lead to less size distortion and power loss than the nonparametric alternative (see Leybourne and McCabe 1994; Perron and Ng 1998).

9. \(\zeta_{it}\) is a function of \(\epsilon_{it}\) and \(\eta_{it}\). Their variances are linked as follows. Defining \(s_i = \sigma_{\epsilon}^2/\sigma_{\eta}^2\), \(\theta_i\) is then related to \(s_i\) by \(s_i = (1 - \theta_i)^2/\theta_i\). We notice that \(s_i = 0\) if and only if \(\theta_i = 1\).
critical, the sensitivity of results to the choice of lag length is checked by choosing \( p_s \) in equation (6) to be larger by one or two than indicated by the best fittings.

Finally, the residual series accounting for serial correlation,

\[
q_{it}^* = q_{it} - \hat{\theta}_t(L)q_{it}
\]

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

2.3 Particular Benefits for PPP Studies

Cross-sectional Dependence. Real exchange rates are known to be highly dependent. This arises for two reasons. First, there is economic dependence between the countries’ price levels and exchange rates. Second, dependence is due to the construction of the real exchange rates with respect to some common benchmark currency. Consequently, any independent variation in the benchmark country’s price level or the value of its currency shows up in all the real exchange rates. O’Connell (1998) shows the importance of accounting for cross-sectional dependence. His simulations indicate that the rejection of the unit root in the panel studies may be due to the failure to account for cross-sectional dependence in real exchange rates; 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 may tend to reject the null of unit root far too frequently, giving false support for PPP. Yet, cross-sectional correlation has been largely ignored because of econometric difficulties in dealing with it. For example, the frequently used LLC test assumes independence of the units.

An important advantage of the NH test is that it accounts for cross-sectional dependence naturally. In notation, this implies that the error covariance matrix, \( \Sigma_e \), need not be diagonal. Thus, the testing procedure captures information regarding correlations among real exchange rates, and consequently has an implication for the test power. When there is no cross-sectional dependence, that is, \( \Sigma_e \) 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.

Invariance to the Benchmark Currency. The NH test is also invariant to the choice of the benchmark currency.\(^{10}\) Many studies report stronger support for PPP (rejec-

---

\(^{10}\) A new panel of real exchange rates benchmarking against a different currency than the U.S. dollar can be constructed by the following linear transformation,

\[
\tilde{q}_t = Aq_t,
\]

where \( A \) is the nonsingular transformation matrix as defined in the appendix of O’Connell (1998). Letting

\[
\tilde{S} = ASA' = T^{-1}\Sigma_c^{1/2}T^{-1}\Sigma_c^{1/2}T = \tilde{C}A^2A' = T^{-2}\sum_{t=1}^T \sum_{i=1}^N (\tilde{q}_i - \tilde{q})' \sum_{i=1}^N (\tilde{q}_i - \tilde{q})
\]

are defined in the text, the invariance property holds simply because

\[
\xi_n = tr[S^{-1}\tilde{C}] = tr[(A^{-1})'S^{-1}A^{-1}(ACA')] = tr[S^{-1}\tilde{C}].
\]
tions of the unit root), when the German mark is used as a numeraire rather than the U.S. dollar. This may be related to the large upward and consequent large downward swing of the dollar in the 1980s that dominates all the dollar-denominated real exchange rates.\textsuperscript{11} This “numeraire effect” seems unjustifiable on theoretical grounds. By construction, a panel of real exchange rates based on another currency is simply a linear transformation from the panel based on the original 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. RESULTS

The simulations are done for three different panels: the full panel of twenty-three industrial countries over 1949–96, twelve European countries for the full sample period and the general floating rate period (1973–96).\textsuperscript{12}

To simulate the small sample distributions of the NH test statistic under the alternative specifications, we proceed as follows.

1. Artificial data series are generated that follow the estimated stationary and nonstationary panel processes as described in section 1.2. The coefficient and the error covariance estimates are those from the SURE estimation, which is done separately for each of the three panels.

2. Each panel is generated five thousand times for the particular dimensions of that panel, for example, the full panel consisting of twenty-three series each forty-eight years long.

3. The NH test statistic is then calculated for each of the five thousand artificial panels. Notice that at this stage, the person applying the test does not know the “true” SURE processes. He just has one of the generated panel series and calculates the NH statistic on that series. In other words, he first fits equation (6) to each series in the generated panel, forms the residual series according to equation (7), and subsequently calculates the NH test for the residual panel.

The distributions of the generated test statistics are plotted in Figures 1 to 3 for the best-fitting stationary and nonstationary process for each of the three panels under study. Notice that these simulated distributions are now specific to the number of observations and the time series properties of our particular series in the panels. For example, they incorporate the estimated covariance structure between the real exchange rates. The distribution of the test statistic conditional on the AR(p) model

\textsuperscript{11} For discussion on the special role of the U.S. dollar in PPP studies, see, for example, Lothian (1998), Papell and Theodoridis (1998) and Koedijk, Schotman, and Van Dijk (1998).

\textsuperscript{12} 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.
is always to the left of the distribution conditional on the nonstationary ARIMA (p,1,0) model.

Finally, the value of the NH statistic is calculated from the original data, and is shown as the dotted line in the figures. By comparing the value of the sample statistic (the dotted line) to the two distributions, we can draw conclusions as to which of the two estimated specifications it is more likely to come from.

Visual inspection of Figures 1 to 3 indicates that the sample statistic clearly comes rather from the stationary distribution. Sensitivity of the simulated distributions to choosing a lag length longer than the true one in equation (6) is also tested by overfitting by one and two additional lags. This does not affect the results. Figure 4 plots the overfitted distributions for the full panel as an example. Overall, the NH test with the parametric estimator of the long-run covariance appears to have excellent power to distinguish the alternatives. Some numerical values describing the power of the NH test to distinguish the processes are given in the upper part of Table 1. These numbers basically tell us nothing more than what can be seen in Figures 1 to 4.

**TABLE 1**

**VARIOUS P-VALUES OF THE TEST STATISTICS FOR THE ESTIMATED PANELS**

<table>
<thead>
<tr>
<th>Statonarity null: NH test</th>
<th>power1</th>
<th>power2</th>
<th>p-value</th>
<th>( \hat{r}_{\text{sample}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parametric correction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EC: 1973–96</td>
<td>0.81</td>
<td>1.00</td>
<td>0.53</td>
<td>2.58</td>
</tr>
<tr>
<td>1-overfit</td>
<td>0.73</td>
<td>1.00</td>
<td>0.20</td>
<td>2.06</td>
</tr>
<tr>
<td>2-overfit</td>
<td>0.67</td>
<td>1.00</td>
<td>0.20</td>
<td>2.08</td>
</tr>
<tr>
<td>EC: 1949–96</td>
<td>0.99</td>
<td>1.00</td>
<td>0.77</td>
<td>4.18</td>
</tr>
<tr>
<td>1-overfit</td>
<td>0.98</td>
<td>1.00</td>
<td>0.83</td>
<td>4.47</td>
</tr>
<tr>
<td>2-overfit</td>
<td>0.97</td>
<td>1.00</td>
<td>0.87</td>
<td>4.81</td>
</tr>
<tr>
<td>FULL: 1949–96</td>
<td>0.99</td>
<td>1.00</td>
<td>0.05</td>
<td>5.84</td>
</tr>
<tr>
<td>1-overfit</td>
<td>0.99</td>
<td>1.00</td>
<td>0.63</td>
<td>6.23</td>
</tr>
<tr>
<td>2-overfit</td>
<td>0.99</td>
<td>1.00</td>
<td>0.65</td>
<td>5.98</td>
</tr>
<tr>
<td>Nonparametric correction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EC: 1973–96</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m = 4 )</td>
<td>0.05</td>
<td>0.17</td>
<td>0.83</td>
<td>2.24</td>
</tr>
<tr>
<td>( m = 8 )</td>
<td>0.05</td>
<td>0.10</td>
<td>0.90</td>
<td>3.22</td>
</tr>
<tr>
<td>( m = 12 )</td>
<td>0.06</td>
<td>0.16</td>
<td>0.86</td>
<td>4.02</td>
</tr>
<tr>
<td>EC: 1949–96</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m = 4 )</td>
<td>0.15</td>
<td>0.58</td>
<td>0.78</td>
<td>1.90</td>
</tr>
<tr>
<td>( m = 8 )</td>
<td>0.07</td>
<td>0.10</td>
<td>0.93</td>
<td>2.11</td>
</tr>
<tr>
<td>( m = 12 )</td>
<td>0.08</td>
<td>0.97</td>
<td>0.05</td>
<td>2.27</td>
</tr>
<tr>
<td>FULL: 1949–96</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m = 4 )</td>
<td>0.07</td>
<td>0.26</td>
<td>0.79</td>
<td>2.91</td>
</tr>
<tr>
<td>( m = 8 )</td>
<td>0.05</td>
<td>0.41</td>
<td>0.59</td>
<td>3.98</td>
</tr>
<tr>
<td>( m = 12 )</td>
<td>0.05</td>
<td>0.58</td>
<td>0.18</td>
<td>4.76</td>
</tr>
<tr>
<td>Unit root null: LLC test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EC: 1973–96</td>
<td>0.61</td>
<td>0.58</td>
<td>0.04</td>
<td>-7.08</td>
</tr>
<tr>
<td>EC: 1949–96</td>
<td>0.78</td>
<td>0.77</td>
<td>0.05</td>
<td>-6.45</td>
</tr>
</tbody>
</table>

**NOTES:**

a. "power1" denotes the probability of rejecting the null at the 5 percent small-sample significance level when the alternative is true. "power2" gives the same probability at the sample statistic \( r_{\text{sample}} \). p-value denotes the percentage of the distribution of the test statistic (under the null) that is to the left of the sample statistic. "1-overfit" and "2-overfit," respectively, denote the cases where the AR orders for each country are chosen to be larger by one or two than the best-fittings would indicate.

b. \( \hat{r}_{\text{sample}} \) are reported in Table 2 for the NH test and in Table 3 for the LLC test.

c. The small-sample critical values of the LLC test are simulated as those for the NH test (see notes to Panel B of Table 2). They at the 5 per-cent level are -6.95 for the EC: 1973–96, and -6.40 for the EC: 1949–96, respectively.
Fig. 1. Distribution of the NH Statistic under the I(0)-Specific and I(1)-Specific Models

Fig. 2. Distribution of the NH Statistic under the I(0)-Specific and I(1)-Specific Models
Fig. 3. Distribution of the NH Statistic under the I(0)-Specific and I(1)-Specific Models

Fig. 4. Distribution of the NH Statistic under the I(0)-Specific and I(1)-Specific Models
"Power1" and "power2" are both close to one, indicating that the probability of rejecting the null when the alternative is true is high, whether measured at the 5 percent small-sample significance level or at the actual sample statistic (the dotted line in the figures). This indicates that we are unlikely to find that PPP holds if the panels were actually generated by the kind of nonstationary process that we estimated in section 1.2.

We now turn to look at the testing problem from the point of view of standard inference. The values of the NH $\xi_N$ statistic are reported in Table 2 (panel A, 1st row) for the full panel of twenty-three industrial countries and the twelve European countries. The asymptotic and small sample bootstrapped critical values for the NH test for our panel dimensions are calculated in Panel B of Table 2. The choice of the small-sample bootstrapped critical values, instead of asymptotic ones, matters for inference: the small-sample critical values give support for PPP across the board, while asymptotic

<table>
<thead>
<tr>
<th>TABLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: NH Test Statistics</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LR cov. estimator</th>
<th>EC12 (N = 12)</th>
<th>FULL PANEL (N = 23)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parametric</td>
<td>4.18***</td>
<td>2.58***</td>
</tr>
<tr>
<td>1-overfit</td>
<td>4.47***</td>
<td>2.06***</td>
</tr>
<tr>
<td>2-overfit</td>
<td>4.81***</td>
<td>2.08***</td>
</tr>
<tr>
<td>Nonparametric</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m = 4$</td>
<td>1.90***</td>
<td>2.24***</td>
</tr>
<tr>
<td>$m = 8$</td>
<td>2.11***</td>
<td>3.22***</td>
</tr>
<tr>
<td>$m = 12$</td>
<td>2.27***</td>
<td>4.02***</td>
</tr>
</tbody>
</table>

**Notes:**

a. "1-overfit" and "2-overfit" denote the values of the test statistic when the AR orders for each country are chosen to be larger by one or two than the best-fittings would indicate.

b. Nonparametric estimates for the long-run covariance are obtained using a Bartlett kernel, $m$ represents the bandwidth number.

c. *** and * denote the cases where the null of PPP cannot be rejected at the 10, 5, and 1 percent level of significance using the small sample critical values in Panel B.

| PANEL B: SMALL SAMPLE AND ASYMPTOTIC CRITICAL VALUES |

<table>
<thead>
<tr>
<th>LR cov. estimator</th>
<th>$T = 48$</th>
<th>$T = 24$</th>
<th>$T = 48$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVS</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>Parametric</td>
<td>4.95</td>
<td>5.35</td>
<td>5.92</td>
</tr>
<tr>
<td>1-overfit</td>
<td>4.93</td>
<td>5.32</td>
<td>5.85</td>
</tr>
<tr>
<td>2-overfit</td>
<td>5.01</td>
<td>5.37</td>
<td>5.78</td>
</tr>
<tr>
<td>Nonparametric</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m = 4$</td>
<td>1.92</td>
<td>1.93</td>
<td>1.95</td>
</tr>
<tr>
<td>$m = 8$</td>
<td>2.09</td>
<td>2.13</td>
<td>2.21</td>
</tr>
<tr>
<td>$m = 12$</td>
<td>2.62</td>
<td>2.65</td>
<td>2.73</td>
</tr>
<tr>
<td>Asymptotics</td>
<td>2.69</td>
<td>2.94</td>
<td>3.51</td>
</tr>
</tbody>
</table>

**Notes:**

a. 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 9, and using the BIC as the selection criterion. Taking these AR orders as given, SUR 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 3,000 time series for each of the three 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 percent of the values of the test-statistics are calculated.

b. Asymptotic critical values are calculated from an approximation to the limiting distributions, CVM(12) and CVM(23) in (4) by truncating the summation at $k = 5000$, and simulating 8000 points from the distribution.
critical values would indicate support for PPP only in the post-73 EC12 panel. Using small-sample critical values, we cannot reject the estimated stationary panel processes at any conventional significance levels. A level as high as 47 percent, 23 percent, and 95 percent is needed for a rejection, respectively, for the shorter EC12, for the longer EC12, and for the whole panel over 1949–96. These results 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.

Overall, the NH test provides strong support for purchasing power parity. The results are all the more remarkable given that the test accounts for cross-sectional correlation and is invariant to the benchmark currency.

Is it appropriate to use the bootstrapped null distribution for inference, or would we be better off using the asymptotic critical values? There is reason to believe that both asymptotic and bootstrapped distributions may overreject the null for the NH test when the AR processes are very persistent. Simulations in Hansen (1999) show that, for persistent autoregressive processes, the right endpoints of the confidence intervals for the autoregressive coefficients are too small. Applying the bootstrapped t-values, instead of the asymptotic ones, reduces the overrejection considerably. Results by Caner and Kilian (1999) reflect the same problem: the extent of size distortions in the KPSS test, when using asymptotic critical values, is shown to increase with persistence, thus leading to faulty rejection of PPP when monthly or quarterly data is used (where the persistence as measured by the AR1 coefficient is typically as high as .94–.99). Since the NH test is a direct multivariate extension of the KPSS test, the univariate findings are likely to apply to it as well. For our annual data, the average persistence as measured by the sum of the AR coefficients is considerably lower, 0.73, 0.84, and 0.87 for the EC 1973–96, EC 1949–96, and the full panel respectively. This should keep us out of the immediate danger zone. Our bootstrapped null distribution of the test statistic is always to the right of the asymptotic distribution, thus being consistent with the expected overrejection when using the asymptotic critical values. Thus, we have reason to believe that the bootstrapped distribution is more appropriate than the asymptotic one.

4. ROBUSTNESS OF RESULTS

4.1 Nonparametric Correction

In the Nyblom-Harvey setup, we can in principle easily check for the sensitivity of inference to alternative estimates of serial correlation and covariance between the real exchange rates. In this section, the parametric adjustment for serial correlation is replaced by nonparametric adjustment. This is done along the line of KPSS (1992) as suggested by Nyblom and Harvey. Nonparametric adjustment does not require us to fit an ARIMA process to each individual series. The estimation of equation (6) can then be skipped, and there is no need to pre-adjust the individual series, as in equation (7), before applying the NH test. Instead, S in equation (3) is replaced by
\[ S(m) = \hat{\Gamma}_0 + \sum_{\tau=1}^{\tau=m} \omega(\tau)[\hat{\Gamma}_\tau + \hat{\Gamma}_\tau'] \]

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 \( \tau \). It is known that the estimated covariance matrix can be sensitive to the choice of the bandwidth parameter \( m \).\(^\text{13}\) The sensitivity of results to different lag truncation points \( m \) is reported in Table 2.

Generally, the test with nonparametric correction has much less small-sample size bias, in contrast to the test with parametric correction (see panel B of Table 2). But the power of the test with the nonparametric correction to distinguish the best-fitting alternatives is relatively poor (see Table 1): the probability of rejecting the PPP null when the panels are nonstationary is clearly below 50 percent in most cases. When we plot (not shown) the simulated distributions of the test statistic under the null and the alternative, they are mostly overlapping, reflecting the lack of power.

Consequently, inference based on such a test with nonparametric correction, has to be interpreted with caution. Nonrejection of the null of stationary real exchange rates by the test here, typically even at 10 percent level of significance for any of the panels when the small-sample critical values are used, cannot be considered to be evidence for PPP.

The poor power performance of the NH test in some cases when the best fittings with a unit root are used as alternatives is consistent with the finding of Engel (2000) for the KPSS test. Both his and our simulations suggest that with nonparametric correction, the test for stationarity null has difficulty in detecting the permanent component, either in the univariate or multivariate context. Our use of parametric correction with the small-sample critical values is therefore a well-justified strategy for inferring on PPP.

We cannot hope to ultimately decide whether any time series processes are I(0) or I(1) in the finite samples, especially having seen the poor power of the nonparamet-

\(^{13}\) There is no way of choosing an optimal value of the parameter for any finite samples. The optimality criteria used gives only 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 a positive semidefinite 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. The optimal bandwidth selection procedure of Andrews and Monahan (1992) leads to power losses when applied to stationarity tests and is therefore not used here (see Lee 1996).
rically corrected NH test. Yet, given that the results with parametric correction are robust to the choice of lag lengths across the panels under investigation, we conclude that there is little evidence against the PPP hypothesis.

4.2 Comparison to Previous Results with the Null of Unit Root

Is the support for PPP specific to the NH testing methodology? Would the conclusions stand with the most frequently used unit root test of Levin-Lin-Chu (1997)? What is 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 on lastly.

Using the LLC test, PPP can be tested in a panel of $N$ countries by testing if $\rho < 0$ in the following ADF regressions:

$$\Delta q_{it} = \rho q_{i,t-1} + \sum_{L=1}^{p} \phi_{iL} \Delta q_{it-L} + \alpha_i + \varepsilon_{it} \quad \text{for } i = 1, \ldots, N \quad (8)$$

where $\alpha_i$ are country-specific intercepts. $\varepsilon_{it}$ 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. Levin, Lin, and Chu (1997) show that the $t$-value of $\hat{\rho}$ in equation (8) under the null has a limiting distribution that is normal.\footnote{This is in contrast to the $t$-value of $\hat{\rho}$ in individual time series, which has a non-normal limiting distribution.}

The following panel, which is practically identical to equation (8), is estimated

$$\Delta \tilde{q}_{it} = \rho \tilde{q}_{i,t-1} + \sum_{L=1}^{p} \phi_{iL} \Delta \tilde{q}_{it-L} + \varepsilon_{it} \quad \text{for } i = 1, \ldots, N \quad (9)$$

where $\tilde{q}_{it} = q_{it} - \frac{1}{T} \sum_{t=1}^{T} q_{it}$. This estimation has been done previously by, for example, 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, that is, the cross-sectionally uncorrelated regression disturbances in equation (8). In the case of real exchange rates, this can be a potentially serious limitation. As suggested by Levin, Lin, and Chu (1997) and Im, Pesaran, and Shin (1997), cross-sectional dependence can be partially accounted for by removing the time-specific effect by time dummies.\footnote{This amounts to subtracting the cross-section averages from the data, thus removing the common component. To do this, we simply deduct the time specific means by replacing $\bar{q}_t$ by $\bar{q}_t = \frac{1}{N} \sum_{i=1}^{N} q_{it}$ in equation (9). This is also done in, for example, Wu (1996) and Coakley and Fuertes (1997).} We can then think of the error term as con-
sisting 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 U.S. dollar clearly is such a common factor for the individual dollar-denominated real exchange rates.

Results with the LLC Test. Table 3 reports the LLC statistics for our panel of twenty-three industrial countries and the twelve EC countries, over 1949–96 and post-1973 separately. 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 the treatment of serial correlation, heteroskedasticity and cross-sectional dependence. It turns out that some of the seemingly contradictory results may indeed be explained by different treatments of these issues in the panel setup.

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 two lagged differences in the estimated model.16 In Table 3, the three rows in each case correspond to \( p = 0,1,2 \). The importance of heteroskedasticity is checked by doing the estimations separately for a model that assumes equal error variance in all countries (\( \text{het} = 0 \) in the Table) and for a model that normalizes according to country-specific error variances (\( \text{het} = 1 \)).

Finally, the effect of cross-sectional dependence due to the common benchmark currency can be seen by comparing the two large columns in Table 3. The results

\[
\Delta q_{it} = \rho q_{i,t-1} + \sum_{i=1}^{n} \Phi_{it} \Delta q_{i,t-1} + \epsilon_{it}
\]

**TABLE 3**

<table>
<thead>
<tr>
<th>Country specific means ( \text{het} = 1 )</th>
<th>Time and country specific means ( \text{het} = 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )</td>
<td>( \hat{\rho} )</td>
</tr>
<tr>
<td>1949–96</td>
<td>23</td>
</tr>
<tr>
<td>Full Panel</td>
<td></td>
</tr>
<tr>
<td>1949–96</td>
<td>12</td>
</tr>
<tr>
<td>EC12</td>
<td></td>
</tr>
<tr>
<td>1973–96</td>
<td>12</td>
</tr>
<tr>
<td>EC12</td>
<td></td>
</tr>
<tr>
<td>1973–96</td>
<td>12</td>
</tr>
<tr>
<td>EC12</td>
<td></td>
</tr>
</tbody>
</table>

**NOTES:**

a. The three numbers reported in a column under each sample refer to estimations with a different number of lagged differences included. The first number corresponds to no lagged differences, the second to \( p = 1 \) and the third to \( p = 2 \).

b. For the panels considered here, the significant rejections of the unit root are indicated by * and ** at the 10, 5, and 1 percent level respectively. For \( N = 23 \), the asymptotic critical values corresponding to the 10, 5, and 1 percent level are \(-7.0\), \(-7.3\) and \(-8.0\), while for \( N = 12 \), they are \(-5.4\), \(-7\) and \(-6.3\).

c. \( \text{het} = 1 \) (\( \text{het} = 0 \)) indicates estimating the model with (without) correction for heteroskedasticity.

d. Country-specific means corresponds to \( \hat{\rho}_t = q_{it} - \frac{1}{T} \sum_{t=1}^{T} q_{i,t} \) and time-specific means further subtracts the time-specific means from the data to arrive at \( \hat{\rho}_t \).

16. Frequently in the literature, an AR(1) or AR(2) model is fitted to the real exchange rates.
when all forms of cross-sectional dependence are ignored are reported under the
column titled "country means" (only country-specific means deducted). The effect of
accounting for the dependence due to the benchmark currency can be seen under the
column titled ‘country and time means’ (with country and time-specific means de-
ducted). Table 3 indicates that the correction for heteroskedasticity is not critical for
inference, while correction for serial correlation is more so. Accounting for cross-
sectional dependence plays some role.

In what follows, we focus on the results with corrections for serial correlation,
heteroskedasticity and cross-sectional dependence. It should be kept in mind that the
correction for cross-sectional dependence is not comparable to that in the NH test,
which accounted for economic dependence as well.

The unit root null is clearly rejected for both of the European country samples
(Table 3), typically at the 1 percent level. In the full panel, results appear to be sen-
tive to the treatment of serial correlation. When no lagged differences are included,
the null is rejected at the 5 percent level, while for \( p = 1 \) or \( p = 2 \) the null can be re-
jected only at about the 20 percent level.\textsuperscript{17}

\textit{Simulation Results.} We turn next to small sample simulations to examine the abil-
ity of the LLC test to distinguish between the alternative best stationary and nonsta-
tionary panel fittings. We focus simulations on the EC 12 panel over both 1949–96
and post-1973. The lower part of Table 1 reports the small sample power of the test.
The experiment is identical to the one done with the NH test previously.

Applying the small sample critical values does not affect the basic conclusion: the
null of unit root can be rejected for the EC12 panel over 1949–96 and 1973–96.
Some summary information on the location of the distributions of the simulated test
statistics under the null and the alternative are reported in Table 1. They indicate that
the LLC test statistic in our sample of EC countries does appear more likely to come
from the stationary distribution, which is thus consistent with the findings with the
NH testing strategy.\textsuperscript{18} This is evident in the bottom of Table 1 where the
\( p \)-values are at the borderline of rejecting the null, and the power of the test seems
to be clearly above 50 percent.

Notice that the simulations are based on the same best-fitting panel processes as
before—only the test applied to the data is different. In other words, the simulated
panel processes incorporate the estimated covariance structure, while the LLC test it-
self accounts for it only partially. This, again, is the setup that the previous studies

\textsuperscript{17} This raises the question as to what value for \( p \) should be used. It is worth noting that 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. In the simulations below,
\( p = 2 \) is being used.

\textsuperscript{18} It is worth noting that correcting for the benchmarking effect reduces the overlap of the distribu-
tions of the test statistic under the alternatives. To permit a comparison, we calculate various \( p \)-values, as
deﬁned in Table 1, for the test statistics without accounting for benchmarking effect. We found that
allowing for benchmarking effect increases “power” of the test by 32 percent (38 percent) and reduces
“\( p \)-value” by 17 percent (12 percent) for the longer (shorter) EC panel. This clearly indicates the increased
power of the test.
applying the LLC test have faced. Our results indicate that the LLC test, even in 
these circumstances, has some power.

Can Our Results be Reconciled with Those of O'Connell (1998)? O'Connell 
found no evidence against the null of unit root in his European panel after applying 
GLS to account for the cross-sectional dependence. A few factors might explain the 
difference.

The most likely explanation in our opinion is that unit root tests of the O'Connell 
type have low power to distinguish the alternatives. The use of GLS not only effec-
tively controls for cross-sectional dependence, but also enhances the power. The gain 
in power, however, may not be enough to tell the alternatives apart, in particular for 
the sample dimensions of concern. O’Connell does not study power performance for 
his actual times series processes. But his simulations with the simplified processes 
(Table 3 in his paper) indicate that the power in the case of his European samples 
(twenty countries over ninety quarters) may well be below 45 percent. This suggests 
much overlapping of the distributions of the test under the alternatives. Nonrejection 
by the GLS unit root test for his European samples could be simply a consequence of 
the test’s lack of power, and thus would have constituted less credible evidence for 
the unit root. On the other hand, use of quarterly data also implies a higher AR(1) co-
efficient making it harder statistically to distinguish the alternative processes. This 
all highlights the potential importance of using the right null hypothesis in the test-
ing—our initial motivation for applying the NH test in this paper.

5. CONCLUDING COMMENTS

The purchasing power parity hypothesis cannot be rejected for our panel of twenty-
three industrial countries, nor for the subsample of twelve European Community 
countries over 1949–96. PPP is found to hold for the European countries in the 
post–Bretton Woods period as well. The application of the Nyblom-Harvey (2000) 
test allowed us to test the null hypothesis directly. Additional complementary evi-
dence in favor of PPP comes from applying the panel version of the conventional unit 
root test. The panel approach adopted here thus points toward the real exchange rates 
of these countries being mean-reverting, despite a great deal of short-term variation.

The findings, moreover, stand when cross-sectional correlation is accounted for. 
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 relationship, has little to do with exchange rate regimes. Indeed, 
our evidence is not any weaker for the longer panel of European countries than for 
the shorter counterpart.

Therefore, the findings with longer panels here, lie between the recent panel evi-
dence under the float and what has been learned from the studies using long time

19. His set of countries is also different from ours. Indeed, one very differently behaving country may 
be all that is needed for the apparent difference. However, the difference in the choice of samples, in com-
parison with the possible low power of the test, appears to be a less important potential explanation.
series. The behavior of real exchange rates, on the basis of this body of evidence, appears to be stable and similar, at least as regards the stability of the first moments. While the 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 continues to serve us well as a first long-run approximation.

LITERATURE CITED


