行政院國家科學委員會專題研究計畫成果報告

檢定工業國家之購買力平價假設與橫斷面依賴

How sure are we about purchasing power parity? Panel evidence with the will of stationary real exchange rates

計畫編號: NSC 88-2415-H-004-010

執行期限: 87 年 8 月 1 日至 89 年 7 月 31 日

主持人: 郭炳伸　　國立政治大學國貿系

E-mail: bsku@nccu.edu.tw

How sure are we about purchasing power parity? Panel evidence with the null of stationary real exchange rates *

Biing-Shen Kuo †
National Chengchi University

and

Anne Mikkola ‡
University of Helsinki, Bank of Finland and CEPR

May, 2000

Abstract

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, i.e. 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).

JEL classification: C22, C52, F31, F47
Keywords: real exchange rate, unit root

*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. We 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

†Associate Professor, Graduate Institute of International Trade, National Chengchi University, Taipei 116, Taiwan; Tel: (886 2) 2939 3091 ext. 81029, Fax: (886 2) 2937 9071, Email: bsku@nccu.edu.tw

‡Post-doctoral Fellow, Research Unit on Economic Structures and Growth, Department of Economics, P.O. Box 54, University of Helsinki, Finland; Tel: (358 9) 191 7984, Fax: (358 9) 191 7980, Email: anne.mikkola@helsinki.fi

1 Introduction

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 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 long-run behavior. 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 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 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 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 amount 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 1990's (see Im, Pesaran and Shin, 1997; Levin, Lin and Chu, 1997, hereafter LLC[1]). For the PPP tests, this has meant testing for the "wrong" null, i.e. the null of the theory not holding. Empirical testing on relatively short univariate time series, e.g. on post-1973 data, typically failed to support PPP. The emerging 'consensus' of the failure of PPP started to shift back towards its acceptance in the 90's. 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 3-5 years (for discussion on the empirical results other than the very latest ones, see Froot and Rogoff (1995) for an excellent survey; and for panel studies, see e.g. 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 non-stationary.

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 100-year UK/US 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 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 practise, 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 non-stationary 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 US/UK real exchange rate series, and in Rudebusch (1993) for the case of the US 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 non-stationarity 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, i.e. 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.[2] 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 US 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, i.e. 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-73 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 700 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 2. The Nyblom-Harvey test and its application to our problem are then discussed in section 3. Results are presented in section 4. Robustness of results to the method of estimation and the choice of the LLC test are discussed in section 5. Section 6 concludes.

2 Best-fitting processes

2.1 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

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

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.[3] Data is available for the following 24 countries over the period 1949-1996: US, 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. Two subpanels are used in the estimation: the 12 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 non-stationary 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.

2.2 Best fitting stationary and non-stationary 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 e.g. Lothian and Taylor, 1996). For the nonstationary specification, ARIMA(r_i,1,0) processes are fitted. They allow for reasonable flexibility in incorporating transitory movements even in the non-stationary 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. The stationary process for country i's real exchange rate is taken to be of the AR(p_i) form

$$q_{it} = \gamma_i + \sum_{j=1}^{p_i} \phi_{ij}\, q_{i,t-j} + v_{it}. \tag{1}$$

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

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

For each real exchange rate series, eq.(1) and eq.(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, i.e. 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 eq.(1) are stacked.

This gives the final panel estimates for $\widehat{\gamma}_i$, $\widehat{\phi}_{ij}$ and the error covariance matrix. The non-stationary panel is estimated analogously by applying SURE to the stacked relationships of the form of eq.(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.[5] The values of the χ^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 24 or 48 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.

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.

3 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.

3.1 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,

$$q_t = \delta + \mu_t + \varepsilon_t,\ \ \varepsilon_t \sim iid(0, \Sigma_\epsilon)$$

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

where

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

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 statistic is

$$\xi_N = tr[S^{-1}C], \tag{3}$$

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

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

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, ξ_N is a test for the null of stationarity. Stationarity will be rejected for large values of ξ_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 ξ_N under the null is

$$\xi_N \to^d \int_0^1 \mathbf{B}(u)'\mathbf{B}(u)du = \sum_{k=1}^{\infty}(\pi k)^{-2}\mathbf{u}_k{}'\mathbf{u}_k, \tag{4}$$