Purchasing Power Parity: The Choice of Price Index*

Maria Cristina T. Terra
EPGE/FGV

Ana Lucia Vahia de Abreu
EPGE/FGV

January 15, 2004

Abstract

Looking closely at the PPP argument, it states that the currencies purchasing power should not change when comparing the same basket goods across countries, and these goods should all be tradable. Hence, if PPP is valid at all, it should be captured by the relative price indices that best fits these two features. We ran a horse race among six different price indices available from the IMF database to see which one would yield higher PPP evidence, and, therefore, better fit the two features. We used RER proxies measured as the ratio of export unit values, wholesale prices, value added deflators, unit labor costs, normalized unit labor costs and consumer prices, for a sample of 16 industrial countries, with quarterly data from 1975 to 2002. PPP was tested using both the ADF and the DF-GLS unit root test of the RER series. The RER measured as WPI ratios was the one for which PPP evidence was found for the larger number of countries: six out of sixteen. The worst measure of all was the RER based on the ratio of foreign CPIs and domestic WPI. No evidence of PPP at all was found for this measure.

JEL Classification Numbers: F31, F41
Keywords: real exchange rate, price indices, purchasing power parity

1 Introduction

The purchasing power parity (PPP) hypothesis, in its original formulation, states that the price levels of two countries should be equal, when measured by the same currency. This is an old idea in economics, but the term was coined only in 1918 by Gustav Cassel. As Cassel (1918) puts it, "(a)s long as anything like free movement of merchandise and a somewhat comprehensive trade between the two countries takes place, the actual rate of exchange cannot deviate very much from this purchasing power parity."

*We thank, without implicating, Luiz Renato Regis de Oliveira Lima and Renato Flôres for useful comments.

Although ever since some variant of PPP has been the building the block for modeling exchange rates long-run behavior, empirical evidence on its validity is, at best, controversial. PPP does not seem to hold in the short run at all, which fits economists assessment that PPP should not hold continuously. However, empirical evidence on long run validity of PPP is also scant. The empirical literature on the subject has investigated possible reasons for the failure of finding hard evidence on long run PPP. Part of the literature credits this failure to the combination of slow speed of convergence, high short run volatility, and not long enough periods of time for testing the long run behavior of the series, for the studies concentrate on post-Bretton-Woods data. The idea is that, with a long enough time span, data on prices and exchange rates would deliver PPP.[1]

Several studies using long span data sets do find more consistent evidence of long-run PPP.[2] The problem with covering a long time frame is that they encompass several different exchange rate regimes. It would be desirable to limit the sample to the pos-Bretton-Woods period. Long time periods are also more prone to include periods with real shock that shift the equilibrium real exchange rate (RER).

Another strand of the literature tries to circumvent the short period of time after Bretton Woods by using panel data. Several such studies reject random walk for the panel. These results, however, solely indicates that random walk is rejected for *at least one* of the RERs used. They do not provide evidence of PPP holding for all of them.[3]

The literature has also turned to nonlinear models to try and explain real exchange rate dynamics. The idea is that transaction costs would yield deviations from PPP, which, in turn, would follow a mean reverting nonlinear process. This would also explain PPP deviations for long periods of time. Sarno and Taylor (2002) present a thorough discussion of this evolving line of research.

An old concern about PPP testing, dating back to Keynes (1932), is the very choice of the price indices to be used. The ideal index should measure the exact same basket of goods in all countries, and these goods should all be tradable. Such an index does not exit, though. The most commonly indices used for testing PPP are Consumer Price Index (CPI) and Wholesale Price Index (WPI). A positive feature of these indices is that they are readily available for most countries and for long time frames. On the negative side, these indices include nontradable goods and they do not measure a common basket of goods across countries. The CPI includes a larger share of nontradable goods than the WPI, hence, one could argue, the WPI would better suit the PPP concept.

This paper revisits this original debate over the price index choice, which should be of an index with the most share of tradable goods and without much variation on the composition of its goods basket across countries. Using PPP testing as a device for spotting those two features, we perform a horse race among six different price indices available from the International Monetary Fund (IMF). We would expect that, if PPP is valid at all, it would be captured

[1] See Froot and Rogoff (1995) and Rogoff (1996).

[2] See Sarno and Taylor (2002) for a brief review of this literature.

[3] Sarno and Taylor (2002) also discusses the results of this literature.

when measuring prices by the price index most in line with those two features. We perform unit root tests for multilateral real exchange rate measures for 16 industrialized countries, for the period from 1975 to 2002. The price indices used are export unit values, wholesale prices, value added deflators, unit labor costs, normalized unit labor costs and consumer prices.

There are studies that test the PPP hypothesis for different price indices such as Dornbusch (1987) that uses CPI, GDP deflator, the GDP deflator for manufacturing and export prices of non-electrical machinery. He finds no evidence of PPP for all price indices studied. Chinn (1998) also implements the PPP testing for different price indices, for several Asian economies. He uses CPI, WPI, PPI and export unit value index. The PPI based results indicate some support for the PPP hypothesis.

Regarding the estimation method, the very early empirical literature tested PPP by estimating simple ordinary least square regressions of price indices on exchange rates. With the evolution of time series econometric modeling, unit root tests based either on augmented Dickey-Fuller (ADF) or variance ratio tests became popular in this literature, along with cointegration studies. Economists have identified the low power of those tests as one possible explanation for the failure to reject random walk from RER series. Sarno and Taylor (2002) perform a Monte Carlo experiment where they simulate data based on a AR(1) model for the RER using different values for the autoregressive coefficient, as estimated in the literature. Using the simulated data, they find that "the probability of rejecting the null hypothesis of a random walk real exchange rate, when, in fact, the real rate is mean reverting, would only be somewhere between about 5 and 7.5 percent." To mitigate this problem, we follow Taylor (2002) and use the Dickey-Fuller test using generalized least squares (DF-GLS) developed by Elliot and al. (1996). This test is a modification of ADF test that increases its power without otherwise altering the method of testing.

Our main results are the following. First, the RER constructed with WPIs supports the PPP hypothesis for the larger number of countries. Hence, this index seems to be the one that best represents tradable goods with a common basket of goods for all countries. Second, when using export unit values, the PPP is verified for only 4 countries. This index includes only goods that are actually traded by the country, hence its goods baskets composition most probably differs across countries to a greater extent, compared to the other indices. Third, unit labor cost ratios are a poor proxy for the relative price of tradable goods. Fourth, for the RER measured as the ratio of foreign CPI and domestic WPI, we find no evidence of PPP holding. Fifth, deterministic trends were found to be significant, possibly indicating some Balassa-Samuelson effect. This is consistent with the idea that CPI has a large share of nontradable goods which are not arbitraged across countries.

The paper is organized as follows. Section two presents the purchasing power parity argument, and its relation to the price indices used to calculate relative purchasing power. The methodology used in the empirical exercises is presented in section 3. Section 4 presents the data and section 5 the empirical results. Finally, section 6 concludes.

2 Purchasing Power Parity

Absolute PPP states that, abstracting from any trade frictions, price levels in two economies should be equal, when measure in the same currency, that is:

$$\frac{EP^*}{P} = 1, \tag{1}$$

where E is the exchange rate, and P and P^* are the price indices in home and foreign countries, respectively. In reality, impediments to trade, such as transport costs and trade barriers, prevent prices to be perfectly equalized. Trade restrictions do not preclude prices from being arbitraged, though, so that prices in different countries should be closely related. Relative PPP allows for obstacles to trade that drive a wedge between the purchasing power of currencies. It states that exchange rate change should reflect relative prices changes:

$$\widehat{E} = \widehat{P} - \widehat{P^*}, \tag{2}$$

where $\widehat{X} = \frac{d \log X}{X}$. Relative PPP should hold when the difference in prices driven by trade frictions do not change over time.

Going from absolute PPP to relative PPP is not only a way of getting around the qualifications arising from trade frictions. It is also a way to solve the problem of prices that are only reported as indices, as opposed to an actual price of a basket of goods. As the price indices are normalized in a base year, even if absolute PPP held, equation (1) would not hold.

PPP, in both its absolute or relative versions, depicts a relation between tradable goods, for these are the goods that are arbitraged by international trade. Hence, the price indices used for testing either equation (1) or equation (2) should contain only tradable goods. Moreover, the price indices to be compared should be composed of the same basket of goods. Unfortunately, no price index has these two features. Price indices available always contain both tradable and nontradable goods, and its goods composition varies, not only across countries, but also over time.

To illustrate the effect on PPP testing of the presence of nontradable goods in the price index and of differences in the price indices composition, let us represent domestic and foreign price indices by a weighted average of tradable and nontradable goods:

$$P = P_N^{\alpha} P_T^{1-\alpha}, \text{ and}$$
$$P^* = P_N^{*\beta} P_T^{*(1-\beta)},$$

where P_N and P_T represent nontradable and tradable goods, respectively, and α and β are the share of nontradable goods in domestic and foreign price indices, respectively. The currency purchasing power for these two price indices, that is, the real exchange rate (RER), equals:

$$\frac{EP^*}{P} = \left(\frac{EP_T^*}{P_T}\right)\left(\frac{P_T^*}{P_N^*}\right)^{\beta}\left(\frac{P_T}{P_N}\right)^{-\alpha},$$

or, in percent changes:

$$\widehat{E}+\widehat{P^{*}}-\widehat{P}=\left(\widehat{E}+\widehat{P_{T}^{*}}-\widehat{P_{T}}\right)-\beta\left(\widehat{P_{T}^{*}}-\widehat{P_{N}^{*}}\right)+\alpha\left(\widehat{P_{T}}-\widehat{P_{N}}\right). \tag{3}$$

International trade arbitrages prices of tradable goods only, so that just the first term in equation (3) should equal zero. PPP failure in empirical testing could be caused by the presence of nontradable goods in the price index. The higher the share of nontradable goods, given by parameters α and β, the higher the impact of nontradable goods relative prices on the currency relative purchasing power.

In addition to the presence of nontradable in the price index, they are also measured differently across countries. This is already partially captured by the difference in parameters α and β. However, the tradable goods composites P_T^* and P_T may also be comprised of different goods basket. Let these indices contain two goods: an exportable and an importable good, with prices P_X and P_M, respectively. The tradables indices in each country may, then, be represented by:

$$P_T = P_X^a P_M^{1-a}, \text{ and}$$
$$P_T^* = P_X^{*b} P_M^{*(1-b)},$$

where a and b are the weights of exportables in each index. Substituting these definitions in equation (3), we get:

$$\begin{aligned}\widehat{E}+\widehat{P^{*}}-\widehat{P} &= b\left(\widehat{E}+\widehat{P_{X}^{*}}-\widehat{P_{X}}\right)+(1-b)\left(\widehat{E}+\widehat{P_{M}^{*}}-\widehat{P_{M}}\right)+ \\ &+(b-a)\left(\widehat{P_{X}}-\widehat{P_{M}}\right)+ \\ &+\beta\left(\widehat{P_{N}^{*}}-\widehat{P_{T}^{*}}\right)-\alpha\left(\widehat{P_{N}}-\widehat{P_{T}}\right).\end{aligned} \tag{4}$$

Now, only the first line in equation (4) would be equal to zero by international price arbitrage. The second line represents changes in measured currency purchasing power due to differences in indices composition. When the indices have the same basket composition we have that $a = b$, and the second line equals zero. The third line captures the effect of the presence of nontradable goods, as discussed above.

Forty years ago Balassa (1964) and Samuelson (1964) set forth the first and most influential model for PPP deviations. They observed that nontradable good price tend to be higher relative to prices of tradable goods in high-income countries compared to low-income countries. Balassa and Samuelson explained this empirical regularity by conjecturing that this relative price differential reflected the fact that richer economies have higher relative productivity in the tradable goods sector. Given competitive pressures within each country for workers with similar skills to receive similar wages in the two sectors, relatively rapid productivity growth in the tradables sector would tend, other things being equal, to push up the relative cost of production in the nontradables sector and, hence, the relative price of nontradables. If international price arbitrage equalizes relative price of tradable goods across countries, such an increase in the relative price of nontradables would, in turn, give rise to an increase in the currency purchasing power for the higher income country, that is a RER appreciation. (See, for instance, Rogoff, 1996, and Isard and Symansky, 1996). In terms of equation (4), taking $\alpha = \beta$ to simplify the argument, the Balassa-Samuelson effect states that, in average, the third line of the equation should be negative when home country is richer than the foreign country.

There is a large literature studying the effect of real variables on deviations of PPP. The RER is modeled as a function of several real variables, such as international terms of trade, trade policy, capital and aid flows, technology and productivity (see, for instance, Baumol and Bowen, 1996, Froot and Rogoff, 1995, De Gregorio, Giovannini and Wolf, 1994, Elbadawi, 1994, and Edwards, 1989, 1994).

3 Methodology

As observed in Rogoff (1996), the empirical literature on PPP has arrived at a consensus on a couple of basic facts. First, a number of recent studies have indicated with fairly persuasive evidence that real exchange rates tend toward purchasing power parity in the long run. Consensus estimates suggest, however, that the speed of convergence to PPP is extremely slow. Second, short-run deviations from PPP are large and volatile. Froot and Rogoff (1995) emphasize that a broad body of evidence suggests that the real exchange rate is not a random walk, and that shocks to the real exchange rate damp out over time, albeit very slowly. They say that, because the convergence to PPP is relatively slow, it is not easy to empirically distinguish between a random walk and a stationary RER that reverts very slowly.

The early empirical literature - until the late 1970s - on PPP testing is based on estimates of an equation in the following form:

$$s_t = \alpha + \beta p_t + \beta^* p_t^* + \omega_t, \tag{5}$$

where s_t is the nominal exchange rate, p_t is the domestic price, p_t^* is the foreign price, all in logs, α, β and β^* are parameters to be estimated, and ω_t is an error term. A test of the restrictions $\alpha = 0$, $\beta = 1$, $\beta^* = -1$ would be interpreted as a test of absolute PPP. In particular, a distinction is often made between the test that β and β^* are equal and of opposite signs - the symmetry condition - and the test that they are equal to $(1, -1)$ - the proportionality condition.

The empirical literature based on the estimation of equation (5) generally suggest rejection of PPP hypothesis. Contrary to other studies, Frenkel (1978) obtains estimates of β and β^* very close to plus and minus unity, on data for high inflation countries, confirming PPP hypothesis. The drawback of that work is that Frenkel does not investigate the residuals stochastic properties. If the residuals are not stationary, the RER is nonstationary, that is, PPP is not valid.

That is, actually, the main problem of this early empirical literature: it does not investigate the stationarity of the residuals. If both nominal exchange rates and relative prices are nonstationary variables and are not cointegrated, then equation (5) is a spurious regression and conventional OLS-based statistical inference is invalid. If the error term in equation (5) is stationary, however, then a strong long run linear relationship exists between nominal exchange rate and relative prices, but conventional statistical inference is still not valid because of the bias present in the estimated standard errors (Sarno and Taylor, 2002).

The next stage in the development of this literature was to analyze the nonstationarity of the RER. When the RER is nonstationary, the series will present a unit root. In this case, the PPP hypothesis is rejected, since it requires the RER to fluctuate around some constant. Evidence against unit root behavior emerges when the RER fluctuates around a fixed mean (constant), with a tendency to return to it. In that case, the effects of shocks will dissipate and the series will revert to its long run mean level. Therefore, if RER is stationary, the PPP can be viewed as a good long run approximation for the RER behavior.

From the mid 1980s onwards, the augmented Dickey-Fuller (ADF) test has been frequently used to test RER stationarity. This test investigates whether the real exchange rate series has stochastic trend. It is based on the estimation of the following equation:

$$(1-L)\,q_t = a + bt + \gamma q_{t-1} + \sum_{j=1}^{p} c_j\,(1-L)\,q_{t-j-1} + \varepsilon_t, \tag{6}$$

where L is the lag operator, $q_t = \log(\mathrm{RER}_t)$, a is the intercept or drift, bt is the linear time trend, p is the number of lags of the RER used in the estimation, and ε_t is the residual. The ADF statistic is the t-statistic for the γ coefficient.

The null hypothesis of the test is $\gamma = 0$ and the alternative hypothesis is $\gamma < 0$. If the test does not reject the null hypothesis, it implies that the RER series presents a unit root. The problem with the ADF test is that it has low power to discriminate between $\gamma = 0$ and a negative value for γ, but very close to zero. For the analysis of PPP, this low power is a problem because, empirically, when the mean reversion occurs ($\gamma < 0$), it does so a very slow speed of convergence, that is, the value of γ is very near zero.

The generalized-least-square (GLS) version of the Dickey Fuller (DF) test suggested by Elliot et al. (1996) has more power that the ADF, being most appropriate for PPP testing. Basically, the test removes means and linear trends from the series, and then performs the unit root test. This modification increases the power of the test without otherwise altering the method of testing. Cheung and Lai (1998) tests PPP for five industrial countries using both the ADF and the DF-GLS tests. They find that the ADF tests verifies stationarity for only two of the ten bilateral RERs studied, whereas the DF-GLS test unravels stationarity in all but two of the series. Taylor (2000) uses the DF-GLS test to investigate PPP for twenty countries, with one hundred years of data.

A Monte Carlo simulation results suggest that the ADF test applied to a demeaned time series or detrended time series, using a data-dependent lag length selection procedure, has the best overall performance in terms of small-sample size and power. Demeaned time series is the case where each series is replaced by the residuals from a regression on a constant and detrended time series is the case where the regression is on constant and a linear trend.

Elliot et al. (1996) define a quasi-difference of q_t that depends on the value a representing the specific point alternative against which we wish to test the null hypotheses, that is:

$$d(q_t|a) = q_t \qquad \text{if } t = 1 \text{ and}$$

$$d(q_t|a) = q_t - aq_{t-1} \qquad \text{if } t > 1.$$

Next consider an OLS regression of the quasi-differenced data $d(q_t|a)$ on the quasi-differenced $d(x_t|a)$:

$$d(q_t|a) = d(x_t|a)'\delta(a) + \eta_t$$

where x_t contains either a constant or a constant and trend, and $\delta(a)$ is the OLS estimate from the regression.

Elliot et al. (1996) demonstrated that for $a = \overline{a}$ the power of the test is maximum:

$$\overline{a} = 1 - \frac{7}{T} \qquad \text{if } x_t = \{1\} \text{ and}$$

$$\overline{a} = 1 - \frac{13,5}{T} \qquad \text{if } x_t = \{1, t\}.$$

where T is the number of sample variables.

The series q_t to be tested is then replaced in the ADF regression by $q_t^d \equiv q_t - x_t'\widehat{\delta}(a)$. Note that since the q_t^d are detrended, we do not include the x_t, or $\alpha + bt$ as in equation (6), in the DF-GLS test equation.

Then the DF-GLS test involves estimation of the following equation:

$$\Delta q_t^d = \gamma q_{t-1}^d + \sum_{j=1}^{p} c_j \Delta q_{t-j-1}^d + \varepsilon_t. \tag{7}$$

As with the ADF test, we consider the t-ratio for $\widehat{\gamma}$ from this test equation. Elliot et al. (1996) simulate the critical values of the test statistic for $T = \{50, 100, 200, \infty\}$.

DF-GLS is the test that will be used for PPP testing in this paper because, first, it is a solution suggested by the literature for the power problem (Taylor, 2000) and, second, it allows deterministic trend in this spirit of the Balassa-Samuelson effect.

4 Data

We use the following price indices from data from the IMF's International Financial Statistics: export unit value, consumer price index (CPI), wholesale price index (WPI), unit labor cost, normalized unit labor cost and relative value added deflator. We use quarterly data for 16 industrialized countries: Austria, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, The Netherlands, Norway, Spain, Sweden, Switzerland, UK and USA. The data for CPI, unit labor cost and normalized unit labor cost ranges from 1975 to 2002. WPI and value added deflator have data from 1975 to 1997, and export unit value from 1975 to 1998.

Export unit value is an indicator for export costs and prices. It is measured as a weighted average of exported goods prices. There are two caveats about this measure. First, this index includes only tradable goods, but not all of them. It includes only goods that are actually exported, but does not compute all potentially exportable goods. It also leaves out imported or importable goods. Second, and a very important caveat that should be emphasized, the basket of goods differs across countries to a greater extent for export unit value than for the other indices. The composition of goods in this index depends on the country's export pattern. As the export pattern differs substantially across countries, so does the composition of the export unit value.

The consumer price indices has a higher share of nontradable goods than the wholesale price indices. One advantage of CPIs is that is available for a larger number of countries and with greater frequency than the other price indices. On the negative side, CPI and WPI includes several factors which may differ across countries, such as price controls, subsidies, indirect taxes and prices of imported goods. These factors may influence the results of PPP testing. Also, CPIs and WPIs are not based on the same basket of goods for different countries, for they reflect different consumption patterns.

Unit labor costs is an indicator for the labor costs, which is an important factor of production in the manufacturing sector. Unit labor costs may be calculated either directly, as total labor costs divided by the total value of output, or indirectly, as the average wage rate divided by labor productivity. This index has the following advantages. First, unit labor costs are defined similarly across industrial countries. Second, as labor costs usually represent the largest share in the total cost of production, the labor cost is a good proxy for production cost. Again, however, there is drawback. The main limitation of the relative unit labor costs as proxy for RER is that they take into account only one factor of production. To the extent that the capital/labor ratio differs across countries, this may introduce a bias into the index.

Normalized unit labor costs is an indicator for the labor costs that removes the distortions arising from cyclical changes in productivity. The advantages of this index is to remove the occasional distortions by cyclical changes in productivity. Productivity changes occur largely due to changes in hours worked that do not correspond closely to changes in the effective inputs of labor. The series on normalized unit labor costs is calculated by dividing labor costs per unit of value added adjusted so as to eliminate the estimated effect of the cyclical swings in economic activity on productivity.

Relative value added deflators is an indicator for the cost (per unit of real value added) of all factors of production in the manufacturing sector. The advantage of this index is that, differently from unit labor costs that take into account only the labor cost, it includes the cost of all factors of production. The main practical disadvantage of value added measures is the lack of cross-country comparability with regard to both concept and commodity composition. Also, they are typically available only for the manufacturing sector, and often with a substantial delay.

We use the multilateral real exchange rate to PPP testing. As stated by Edwards (1989), in a world where the main currencies are floating there are many different bilateral rates, and there is no reason why one rate should be preferred over another. For this reason, indices of RER that take into account the behavior of all the relevant bilateral exchange rate were considered.

Following the methodology of IMF, the RER was computed as:

$$RER_{\,i} = \prod_{j \neq i} \left[\frac{E_i P_{\,i}}{E_j P_{\,j}} \right]^{W_{ij}} \tag{8}$$

where the nominal exchange rate is period-average US dollars per unit of national currency and W_{ij} is the weight[4] attached by country i to country j.

The IMF's International Financial Statistics presents the computed RER, as in equation (8), for all indices. The only RER we computed with original price indices and nominal exchange rates from the IMF was the RER measured as the ratio of foreign countries' WPI over domestic country's CPI.

5 Empirical Results

We now present the results of PPP testing for the seven different proxies for RER: ratios of export unit values, CPIs, WPIs, unit labor costs, normalized unit labor costs,relative value added deflators, and the ratio between WPI and CPI. We tested PPP for each one of the indices, for each country, using both the traditional augmented Dickey-Fuller test and the power-enhancing Dickey-Fuller test using generalized least squares estimation.

We start with PPP testing for the RER based on export unit values. The results of the ADF unit root tests are presented in Table 1. The unit root null hypothesis cannot be rejected in all but two countries: France and Sweden. When we allow for a trend, unit root is rejected only for Switzerland. A simple OLS regression on a constant and a trend indicates the presence of a trend for Canada and Switzerland. Hence, the results of both detrended ADF and simple OLS indicate that, for Switzerland, the RER based on export unit values has a deterministic trend, although the trend component amounts to only 0.04% per quarter. Nonetheless, we could not reject random walk for this series in the

[4] For a discussion about the computation of weights (W_{ij}), see the appendix.

estimation without trend, that is, in the "demeaned" result. As Taylor (2000) puts it, "it is necessary to allow for slowly-evolving deterministic trends. As an empirical matter, they are usually found to be "small". However, their omission would undoubtedly upset any study of the deviations of real exchange rates over the very long run".

The results for the DF-GLS test are presented in Table 2. Differently from the ADF test, the DF-GLS test rejects the unit root null for Sweden. Nevertheless, with the DF-GLS test there are four countries, instead of only two, for which the unit root can be rejected: France, Germany, Italy and the Netherlands. The detrended Switzerland RER series also does not present a unit root, and so does the detrended France series. Comparing the two tests, the DF-GLS captures convergence in a larger number of countries compared to the ADF test, as expected. Yet, we could not reject the present of unit roots in most of the series, in both tests.

Even though the export unit values index only includes tradable goods, the PPP hypothesis is valid for only, at most, four countries out of sixteen. The reason for this result may be that the goods basket composition differs substantially across countries. When comparing export unit values for two countries, we are comparing the weighted values for two different baskets of goods. Hence, even if the traded goods prices are arbitraged by trade, the value of the index could follow different paths in different countries due to the difference in the index composition in each of the countries.

For the RER series based on wholesale price indices, the ADF tests does not reject the unit root null for any of the series, as shown in Table 3. As presented in Table 4, using the more powerful DF-GLS, unit root is rejected for six countries: Finland, France, Germany, Italy, Switzerland and Spain. For the detrended estimation, unit root is not rejected for any of the countries. As we will see, this is the RER series for which PPP is valid for a larger set of countries.

Tables 5 and 6 present the results of the ADF and DF-GLS tests, respectively, for the RER series constructed as CPI ratios. The presence of unit root cannot be rejected for any of the countries, using the ADF test. Using the DF-GLS test, four countries, Denmark, Finland, Italy and Norway, are found not to present unit root in their RER series. The result for Switzerland RER series is analogous to the one for its RER series based on export unit values: we cannot reject the unit root null for its demeaned series, but, once a trend is included, the series becomes stationary. This result indicates that there is a also deterministic trend in the RER based of CPI ratios, and this is the reason for the non validity of PPP hypothesis.

The CPIs is more heavily weighted with nontradable goods than tradable goods, when compared with WPIs. As shown in equation (**??**), the higher the weight of nontradable goods in the price index composition, the larger may potentially be the deviations from PPP. That seems to be the case for France, Germany, Spain and Switzerland. Their RER series based on WPI were stationary, but the ones based on CPI presented unit roots. The odd cases are Denmark and Norway, for their RER series present unit roots when based on

WPIs, but not when constructed using CPIs.

The results of PPP testing for RER based on unit labor cost and on normalized unit labor cost are very similar. The ADF test does not detect stationarity for any of the two series, as shown in Tables 7 and 9. Adding a trend to the estimation results in rejection of unit root for France for the two series, and for Sweden for the unit labor cost series. The estimation with DF-GLS somewhat improves the results. Table 8 shows that the unit root null is rejected for Denmark, Italy and Sweden, for the RER based on unit labor cost. For the normalizes series, unit root is rejected only for Canada and Denmark, as presented in Table 10. We cannot reject unit roots for any of the detrended estimation, for the two sets of RER series. This means that no deterministic trend explain the unit root evidence.

These results indicates that the RER proxied by the ratio of unit labor cost, normalized or not, is a poor proxy for the relative prices of tradable goods. One possible explanation is the fact the capital to labor ratio differs substantially across countries, so that the labor cost becomes a poor reflection of relative prices.

The results for the value-added-RER series are interesting. The results from ADF, in Table 11, detects no unit root only for Switzerland. The DF-GLS, on the other hand, rejects the unit root null for five countries: France, Germany, Spain, Sweden and Switzerland. These results, in Table 12, are close to the ones for the RER series based on WPI, for which stationarity was found for six countries.

The worst results are those for the RER measures as a ratio of foreign countries CPI and domestic country WPI. No evidence of stationarity of theses series were found, using both the ADF and the DF-GLS unit root tests. The results are presented in Tables 13 and 14. This proxy for RER suffers from two of the problems that could causes PPP deviations, as detected in equation (**??**): some of the price indices have a large share of nontradable goods (the CPIs), and the composition of foreign and domestic indices are substantially different (as we a using simultaneously CPIs and WPIs)

Table 15 presents a summary of the results. It shows the countries for which we found evidence of PPP for each RER proxy and unit root test used. The first striking result is that PPP is detected in a much larger set of countries when we use the DF-GLS test, compared to the ADF test. This result was expected. The DF-GLS has more power than the ADF, so that it is more competent to reject the unit root null when the speed of convergence is low.

The RER proxy leader in stationarity is the one constructed as WPIs ratios, presenting PPP evidence for six of the sixteen countries studied. This is a signal that this price index is the one that better fits the requirement for PPP: more uniform goods composition across countries and low share of nontradable goods. The second place goes to the RER based on value added. PPP evidence was found for five of the countries, for this RER proxy. The third position is a draw between the RER based on export unit values and the one based on CPIs ratio: they both yield PPP for four of the countries studied. Unquestionably, the very last place goes to the RER constructed as the ratio between foreign countries CPIs and domestic country WPI, as no PPP evidence was found for that measure.

Looking at the countries' perspective, France is the country for which PPP evidence was found in the larger number of RER series. There is some evidence of PPP for France for five of the seven RER proxies used. Switzerland and Italy follow closely, with PPP evidence for four of the RER series. No evidence of PPP was found in any of the series for five countries: Austria, Belgium, Japan, United Kingdom and United States.

6 Concluding Remarks

There is a huge literature testing the PPP hypothesis, most of it using either CPIs or WPIs ratios as proxies of relative currencies purchasing power, that is, of the RER. Looking closely at the PPP argument, it states that the currencies purchasing power should not change when comparing the same basket goods across countries, and these goods should all be tradable. Neither of those price indices used in PPP testing fully satisfy these two criteria: they include nontradable goods and their basket composition differs across countries. We observe that, if PPP is valid at all, it should be captured by the relative price indices that best fits these two features. Hence, we ran a horse race among six different price indices available from the IMF database to see which one would yield higher PPP evidence. We used RER proxies measured as the ratio of export unit values, wholesale prices, value added deflators, unit labor costs, normalized unit labor costs and consumer prices. PPP was tested using both the ADF and the DF-GLS unit root test of the RER series.

The RER measured as WPI ratios was the one for which PPP evidence was found for the larger number of countries: six out of sixteen. This is an indication that, from all indices used, WPI seems to be the one with larger composition of tradable goods and with least variation in its goods basket composition across countries.

The second best RER measure was the ratio of export unit values. On the one hand, this is an index composed solely of tradable goods, so it surely fits one of the criteria. On the other hand, the index composition may vary substantially across countries, as the export pattern does differ a lot across countries.

Unit labor costs and normalized unit labor cost proved to be poor measures of tradable goods, as PPP evidence was found for a small number of countries, when RER was measured by them. However, the worst measure of all was the RER based on the ratio of foreign CPI and domestic WPI. No evidence of PPP at all was found for this measure.

Finally, deterministic trends were found to be significant in several cases, possibly indicating some Balassa-Samuelson effect.

Overall, this paper also identifies the importance of the price index choice to compute the RER for PPP testing. The results differs substantially when different proxies for the RER were used. Nevertheless, some consistency was present. We found PPP evidence for RER series for France, Switzerland and

Italy for most of the RER proxies, whereas no PPP evidence was found for Austria, Belgium, Japan, United Kingdom and United States for any of the proxies.

7 Appendix

This appendix presents the methodologies for computation of the weights (W_{ij}), published in the Fund's International Financial Statistics.

We begin with the methodology for the weights used in the computation of RER based on relative export unit values, wholesale prices, value added deflators, unit labor costs and normalized unit labor costs.

From January 1991 onwards, W_{ij} uses data on trade and consumption of manufactured goods over the period 1989-91. Before that, the weights used in the computation of RER were based on 1980 data.

Let there be k markets in which the producers of country i and country j compete. Let T_l^k represent the sales of country l in market k. Let s_j^k be country $j\prime s$ market share in market k and w_i^k be the share of country $i\prime s$ output sold in market k, which is to say:

$$s_j^k = \frac{T_j^k}{\sum\limits_l T_l^k} \text{ and} \tag{9}$$

$$w_i^k = \frac{T_i^k}{\sum\limits_n T_i^n}. \tag{10}$$

Then, the weight attached to country j by country i is:

$$W_{ij} = \frac{\sum\limits_k w_i^k s_j^k}{\sum\limits_k w_i^k (1 - s_i^k)}. \tag{11}$$

This weight can be interpreted as the sum over all markets of a gauge of the degree of competition between producers of countries i and j divided by the sum over all markets of a gauge of the degree of competition between producers of country i and all other producers. The world is divided into 22 markets, the first 21 markets being the countries[5] for which RER were being computed by IMF and the last market is called "Rest-of-the-World".

Next, we will present the second methodology that describes the weights used in the computation of RER based on consumer price index.

From January 1990 onwards, W_{ij} is weighted by a set of weights based on trade in manufactures, non-oil primary commodities and, for a set of 46 countries

[5] These 21 countries are Australia, Austria, Belgium, Canada, denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom and the United States.

and regions[6] in which services accounted to meet more than 20 percent of all exports in 1989-90, tourism services covering the three-year period 1988-1990. Prior to January 1990, the weights are for the three-year span 1980-82.

These weights are then aggregated to derive the overall weight attached by country i to country j, W_{ij}. Specifically:

$$W_{ij} = \alpha_i(M)W_{ij}(M) + \alpha_i(P)W_{ij}(P) + \alpha_i(T)W_{ij}(T), \tag{12}$$

where $W_{ij}(M)$, $W_{ij}(P)$ and $W_{ij}(T)$ are weights based on trade in manufactures, primary commodities and tourism services. The factors $\alpha_i(M)$, $\alpha_i(P)$ and $\alpha_i(T)$ are the shares of trade in manufactures, primary commodities and tourism services in country $i\prime s$ external trade, with external computed as the sum of trade in manufactures, primary commodities and tourism services. Observe that $\alpha_i(T) = 0$ for a set of countries in which services accounted to meet less than 20 percent of all exports in 1989-90. For these countries, $\alpha_i(M)$ and $\alpha_i(P)$ are the shares of trade in manufactures and primary commodities in total trade, with total trade being computed as the sum of trade in these two categories.

The weights based on trade in manufactures, $W_{ij}(M)$, and on trade in tourism, $W_{ij}(T)$, are computed in a manner analogous to equation (11). These weights are a weighted sum of a weight reflecting competition in the domestic market, a weight reflecting competition abroad against domestic producers and a weight reflecting competition abroad against exporters.

The weights based on trade in primary commodities, $W_{ij}(P)$, are computed in a very different way. Contrary to manufactured goods and tourism services, primary commodities are assumed to be homogeneous goods. Then, for each commodity, the weight attached to country j by any country should reflect the importance of country j as either a seller or a buyer in the world market. Therefore, for country i, the weight attached to country $j, W_{ij}(P)$, should be a (normalized) sum over all commodity markets of the product of the individual weight of country j in each market h times the importance of market h for country i.

References

[1] Asea, Patrick K. and Enrique G. Mendoza, 1994, “The Balassa-Samuelson Model: A General-Equilibrium Appraisal,” *Review of International Economics* 2 (October).

[6] These 46 countries and regions are Antigua and Borbuda, Austria, The Bahamas, Barbados, Belize, Costa Rica, Cyprus, Dominica, Dominican Republic, Egypt, El salvador, Fijii, France, Germany, Greece, Grenada, Italy, Jamaica, Japan, Jordan Kenya, Kiribati, Maldives, Mali, Malta, Mauritius, Morocco, Nepal, Netherlands Antilles, Portugal, St. Kitts and Nevis, St. Lucia, St. Vincent and the Grenadines, Seychelles, Spain, Syrian Arab Republic, Thailand, Togo, Tunisia, Turkey, United Kingdom, United States, Vanuatu, Western Samoa and Republic of Yemen.

[2] Balassa, Bela, 1964, "The Purchasing Power Parity Doctrine: A Reappraisal," *Journal of Political Economy* 72: 584-96.

[3] Baumgarten, Monica T. P., 1996, "Modelos de Taxa de Câmbio Real de Equilíbrio: uma Aplicação para o Brasil," Dissertação de Mestrado, Departamento de Economia, PUC-Rio.

[4] Breuer, Janice B., 1994, "An Assessment of the Evidence on Purchasing Power Parity," in John Williamson (ed.) *Estimating Equilibrium Exchange Rates* (Washington: Institute for International Economics).

[5] Cassel, Gustav, 1918, "Abnormal Deviations in International Exchanges," *The Economic Journal* 28 (112): 413-415.

[6] Clark, Peter B., Leonardo Bartolini, Tamim Bayomi and Steve Symansky, 1994, "Exchange Rates and Economic Fundamentals: A Framework for Analysis," IMF Occasional Paper No. 115.

[7] Clark, Peter B., Tamim Bayomi, Steve Symansky and Mark Taylor, 1994, "The Robustness of Equilibrium Exchange Rate Calculations to Alternative Assumptions and Methodologies," in John Williamson (ed.) *Estimating Equilibrium Exchange Rates* (Washington: Institute for International Economics).

[8] Cheung, Yin-Wong and Kon S. Lai, 1998, "Parity Reversion in Real Exchange Rates during the Post-Bretton Woods Period," *Journal of International Money and Finance* 17(4): 597-614.

[9] Chinn, Menzie, 1998, "Before the Fall: Were East Asian Currencies Overvalued", NBER Working Paper Series No. 6491.

[10] Chinn, Menzie and Eswar S. Prasad, 2000, "Medium-Term Determinants of Current Accounts in Industrial and Developing Countries: an Empirical Exploration," IMF Working Paper WP/00/46.

[11] De Gregorio, José, Alberto Giovannini and Holger C. Wolf, 1994, "International Evidence on Tradables and Nontradables Inflation," *European Economic Review* 38(June).

[12] De Gregorio, José and Holger C. Wolf, 1994, "Terms of Trade, Productivity, and The Real Exchange Rate", NBER Working Paper Series No. 4807.

[13] Desruelle, Dominique and Alessandro Zanello, 1997, "A Primer on the IMF´s Information Notice System," IMF Working Paper WP/97/71.

[14] Dornbusch, Rudiger, 1987, "Purchasing Power Parity," in John Eatwell, Murray Milgate and Peter Newman (eds.) *The New Palgrave: A Dictionary of Economics*, vol. 3: 1075-85 (New York: Stockton Press, 3rd ed.).

[15] Edwards, Sebastian, 1989, *Real Exchange Rates, Devaluation and Adjustment: Exchange Rate Policy in Developing Countries* (Cambridge, Massachusetts: MIT Press).

[16] Edwards, Sebastian, 1994, "Real and Monetary Determinants of Real Exchange Rate Behavior: Theory and Evidence from Developing Countries," in John Williamson (ed.) *Estimating Equilibrium Exchange Rates* (Washington: Institute for International Economics).

[17] Elbadawi, Ibrahim, 1994, "Estimating Long-Run Equilibrium Real Exchange Rates," in John Williamson (ed.) *Estimating Equilibrium Exchange Rates* (Washington: Institute for International Economics).

[18] Elliott, Graham, Thomas J. Rothenberg and James H. Stock, 1996, "Efficient Tests for an Autoregressive Unit Root," *Econometrica* 64(4): 813-36.

[19] Frenkel, Jacob A., 1978, "Purchasing Power Parity: Doctrinal Perspective and evidence from 1920s", *Journal of International Economics* 8: 169-91.

[20] Froot, Kenneth A. and Kenneth Rogoff, 1991, "Government consumption and the real exchange rate: The empirical evidence," mimeo, Harvard Business School.

[21] Froot, Kenneth A. and Kenneth Rogoff, 1995, "Perspectives on PPP and Long Run Real Exchange Rates," em Gene M. Grossman and Kenneth Rogoff (eds.) *Handbook of International Economics*, vol. 3 (Amsterdam: North-Holland).

[22] Guerguil, Martine and Martin Kaufman, 1998, "Competitiveness and the Evolution of the Real Exchange Rate in Chile", IMF Working Paper WP/98/58.

[23] Hinkle, Lawrence and Peter Montiel, 1999, Exchange Rate Misalignment, A World Bank Research Publication (Oxford University Press). *???É um livro???*

[24] International Monetary Fund, 1984, "Issues in the Assessment of the Exchange Rates of Industrial Countries", IMF Occasional Paper No. 29.

[25] Isard, Peter, 1995, *Exchange Rate Economics*, Cambridge Surveys of Economic Literature (Cambridge, England: Cambridge University Press).

[26] Isard, Peter and Steven Symansky, 1996, "Long-Run Movements in Real Exchange Rates," in Takatoshi Ito, Peter Isard, Steven Symansky and Tamim Bayoumi (eds.) *Exchange Rate Movements and Their Impact in Trade and Investment in the APEC Region*, IMF Occasional Paper No. 145. *???É um livro???*

[27] Isard, Peter and Hamid Faruqee, 1998, "Exchange Rate Assessment: Extensions of the Macroeconomic Balance Approach", IMF Occasional Paper No. 167.

[28] Isard, Peter, Hamid Faruqee, G. Russell Kincaid and Martin Fetherston, 2001, "Methodology for Current Account and Exchange Rate Assessments," IMF Occasional Paper No. 209

[29] Keynes, John M., 1932, *Essays in Persuasion* (New York: Hartcourt Brace).

[30] Kuo, Biing-Shen and Mikkola, Anne, 2001, "How Sure Are We about Purchasing Power Parity ? Panel Evidence with the Null of Stationary Real Exchange Rates," *Journal of Money, Credit and Banking*, August.

[31] Lipschitz, Leslie and Donough McDonald, 1991, "Real Exchange Rates and Competitiveness: A Clarification of Concepts, and Some Measurements for Europe," IMF Working Paper WP/91/25 (Washington: International Monetary Fund).

[32] Marsh, Ian and Stephen Tokarick, 1994, "Competitiveness Indicators: A Theoretical and Empirical Assessment" IMF Working Paper WP/94/29 (Washington: International Monetary Fund).

[33] Rogoff, Kenneth, 1996, "The Purchasing Power Parity Puzzle," *Journal of Economic Literature* 34: 647-68.

[34] Samuelson, Paul A., 1964, "Theoretical Notes on Trade Problems", *Review of Economics and Statistics* 46: 145-54.

[35] Sarno, Lucio and Taylor, Mark P., 2002, "Purchasing Power Parity and the Real Exchange Rate", IMF Staff Papers vol. 49, No 1 (Washington: International Monetary Fund).

[36] Taylor, Alan N., 2002, "A Century of Purchasing-Power Parity," *Review of Economics and Statistics* 84(1): 139-50.

[37] Wickham, Peter, 1987, "A Revised Weighting Scheme for Indicators of Effective Exchange Rates," IMF Working Paper WP/87/87 (Washington: International Monetary Fund).

[38] Williamson, John, 1994, "Estimates of REERs", in John Williamson (ed.) *Estimating Equilibrium Exchange Rates* (Washington: Institute for International Economics).

Table 1: ADF test: RER based on export unit value

	Demeaned		Detrended		OLS
Country	ADF	Lags	ADF	Lags	Trend
Austria	-2.12	4	-2.11	4	0.0000
Belgium	-2.55	0	-2.20	0	0.0000
Canada	-0.31	0	-2.40	0	-0.0002**
Denmark	-1.20	0	-1.52	0	0.0000
Finland	-2.19	0	-2.14	0	0.0000
France	-2.61*	0	-2.90	0	0.0000
Germany	-1.77	0	-1.93	0	0.0000
Italy	-2.41	0	-2.43	0	0.0000
Japan	-1.21	0	-1.74	0	0.0001
Netherlands	-2.42	0	-2.54	0	0.0000
Norway	-2.49	1	-2.79	1	-0.0001
Spain	-1.30	0	-2.15	0	0.0001*
Sweden	-2.69*	0	-2.43	0	0.0000
Switzerland	-1.06	0	-3.40*	0	0.0004***
UK	-2.13	0	-2.64	0	0.0001
US	-1.51	0	-1.50	0	0.0000

Notes: Data from 1975:1 to 1998:2. The lag length is selected by Modified SIC, with maximal lag length equal to 6. Asterisks denotes significance at the *10%, ** 5% , and *** 1% levels. The critical values corresponding to these significance levels are (-2.58, -2.89, -3.51) for the demeaned series and (-3.16, -3.46, -4.06) for the detrended series, respectively.

Table 2: DF-GLS test: RER based on export unit value

	Demeaned		Detrended	
Country	DF-GLS	Lags	DF-GLS	Lags
Austria	-0.34	4	-1.44	4
Belgium	-0.94	0	-1.58	0
Canada	1.53	0	-2.46	0
Denmark	-1.22	0	-1.32	0
Finland	-0.68	0	-1.74	0
France	-2.60***	0	-2.87*	0
Germany	-1.75*	0	-2.02	0
Italy	-2.36**	0	-2.46	0
Japan	-1.21	0	-1.48	0
Netherlands	-2.04**	0	-2.23	0
Norway	-0.94	1	-2.43	1
Spain	-1.25	0	-1.79	0
Sweden	-1.06	0	-1.76	0
Switzerland	-0.12	0	-3.32**	0
UK	-0.81	0	-2.48	0
US	-0.69	0	-1.19	0

Notes: Data from 1975:1 to 1998:2. The lag length is selected by Modified SIC, with maximal lag length equal to 6. Asterisks denotes significance at the *10%, ** 5% , and *** 1% levels. The critical values corresponding to these significance levels are (-1.61, -1.94, -2.59) for the demeaned seriesand (-2.77, -3.07, -3.62) for the detrended series, respectively.

Table 3: ADF test: RER based on wholesale price index

	Demeaned		Detrended		OLS
Country	ADF	Lags	ADF	Lags	Trend
Austria	-1.58	2	-2.63	2	-0.0001**
Belgium	-1.38	1	-1.10	0	0.0000
Canada	-1.88	0	-1.77	0	0.0000
Denmark	-1.51	0	-1.83	0	0.0000
Finland	-1.30	0	-1.46	0	0.0000
France	-1.83	0	-1.92	0	0.0000
Germany	-1.69	0	-2.08	0	0.0000
Italy	-1.90	0	-1.92	0	0.0000
Japan	-1.57	0	-1.71	0	0.0001
Netherlands	-1.39	0	-1.54	0	0.0000
Norway	-1.51	0	-0.58	0	0.0000
Spain	-2.21	0	-2.16	0	0.0000
Sweden	-2.30	0	-2.23	0	0.0000
Switzerland	-2.47	0	-2.70	0	0.0001
UK	-1.77	0	-2.22	0	0.0001
US	-1.12	0	-1.57	0	-0.0001

Notes: Data from 1975:1 to 1997:1. The lag length is selected by Modified SIC, with maximal lag length equal to 6. Asterisks denotes significance at the *10%, ** 5% , and *** 1% levels. The critical values corresponding to these significance levels are (-2.58, -2.90, -3.51) for the demeaned series and (-3.16, -3.46, -4.07) for the detrended series, respectively.

Table 4: DF-GLS test: RER based on wholesale price index

	Demeaned		Detrended	
Country	DF-GLS	Lags	DF-GLS	Lags
Austria	-1.51	2	-2.37	2
Belgium	-0.02	1	-1.08	0
Canada	-1.04	0	-1.56	0
Denmark	-1.52	0	-1.65	0
Finland	-1.71*	1	-1.50	0
France	-1.74*	0	-1.80	0
Germany	-1.69*	0	-1.86	0
Italy	-1.89*	0	-1.95	0
Japan	-1.59	1	-1.86	0
Netherlands	-1.39	0	-1.45	0
Norway	-0.52	0	-0.92	0
Spain	-2.21**	0	-2.22	0
Sweden	-1.52	0	-2.03	0
Switzerland	-2.49**	0	-2.65	0
UK	-0.50	0	-2.09	0
US	-1.22	0	-1.40	0

Notes: Data from 1975:1 to 1997:1. The lag length is selected by Modified SIC, with maximal lag length equal to 6. Asterisks denotes significance at the *10%, ** 5% , and *** 1% levels. The critical values corresponding to these significance levels are (-1.61, -1.94, -2.59) for the demeaned seriesand (-2.78, -3.07, -3.63) for the detrended series, respectively.

Table 5: ADF test: RER based on consumer price index

	Demeaned		Detrended		OLS
Country	ADF	Lags	ADF	Lags	Trend
Austria	-1.52	0	-1.59	0	0.0000
Belgium	-1.63	1	-1.76	1	0.0000
Canada	-1.13	1	-1.58	0	-0.0001
Denmark	-1.85	0	-2.13	0	0.0000
Finland	-1.64	1	-1.55	0	0.0000
France	-1.63	0	-2.26	0	0.0000
Germany	-2.00	0	-1.94	0	0.0000
Italy	-2.00	1	-1.47	0	0.0000
Japan	-2.08	1	-1.79	0	0.0001
Netherlands	-1.76	0	-1.73	0	0.0000
Norway	-1.97	0	-1.65	0	0.0000
Spain	-1.84	0	-1.73	0	0.0000
Sweden	-1.10	0	-2.12	0	-0.0001*
Switzerland	-2.12	0	-2.89	0	0.0001*
UK	-1.69	0	-1.95	0	0.0001
US	-1.04	0	-1.08	0	0.0000

Notes: Data from 1975:1 to 2002:3. The lag length is selected by Modified SIC, with maximal lag length equal to 6. Asterisks denotes significance at the *10%, ** 5% , and *** 1% levels. The critical values corresponding to these significance levels are (-2.58, -2.89, -3.49) for the demeaned series and (-3.15, -3.45, -4.04) for the detrended series, respectively.

Table 6: DF-GLS test: RER based on consumer price index

	Demeaned		Detrended	
Country	DF-GLS	Lags	DF-GLS	Lags
Austria	-0.74	0	-1.66	0
Belgium	-0.90	1	-1.75	1
Canada	-0.20	1	-1.59	0
Denmark	-1.81*	0	-1.90	0
Finland	-1.64*	1	-1.37	0
France	-0.61	0	-2.25	0
Germany	-0.81	1	-1.54	0
Italy	-1.98**	1	-2.07	1
Japan	-0.71	1	-1.72	0
Netherlands	-1.16	0	-1.82	0
Norway	-1.64*	0	-1.93	0
Spain	-1.22	0	-1.62	0
Sweden	0.01	0	-2.16	0
Switzerland	-1.45	0	-2.91*	0
UK	-1.09	0	-1.97	0
US	-0.92	0	-1.16	0

Notes: Data from 1975:1 to 2002:3. The lag length is selected by Modified SIC, with maximal lag length equal to 6. Asterisks denotes significance at the *10%, ** 5% , and *** 1% levels. The critical values corresponding to these significance levels are (-1.61, -1.94, -2.59) for the demeaned seriesand (-2.73, -3.02, -3.57) for the detrended series, respectively.

Table 7: ADF test: RER based on unit labor cost

	Demeaned		Detrended		OLS
Country	ADF	Lags	ADF	Lags	Trend
Austria	-0.30	1	-2.81	1	-0.0002***
Belgium	-1.51	0	-1.09	0	0.0000
Canada	-1.89	0	-1.83	0	0.0000
Denmark	-2.19	2	-2.29	2	0.0000
Finland	-1.75	0	-4.48	0	-0.0004***
France	-0.78	0	-3.29*	0	-0.0002***
Germany	-1.38	0	-2.12	0	0.0001*
Italy	-2.40	0	-2.35	0	0.0000
Japan	-1.89	0	-2.29	0	0.0000
Netherlands	-1.75	1	-1.29	1	0.0000
Norway	0.18	0	-0.27	1	0.0000
Spain	-2.54	0	-2.39	0	0.0000
Sweden	-1.61	0	-3.34*	0	-0.0002***
Switzerland	-0.51	2	-2.62	2	0.0002***
UK	-1.44	0	-1.64	0	0.0000
US	-1.26	0	-1.54	0	-0.0001

Notes: Data from 1975:1 to 2002:3. The lag length is selected by Modified SIC, with maximal lag length equal to 6. Asterisks denotes significance at the *10%, ** 5% , and *** 1% levels. The critical values corresponding to these significance levels are (-2.58, -2.89, -3.49) for the demeaned series and (-3.15, -3.45, -4.04) for the detrended series, respectively.

Table 8: DF-GLS test: RER based on unit labor cost

	Demeaned		Detrended	
Country	DF-GLS	Lags	DF-GLS	Lags
Austria	-0.38	1	-1.26	1
Belgium	0.34	0	-0.97	0
Canada	-1.49	0	-1.60	0
Denmark	-2.21**	2	-2.27	2
Finland	-0.85	0	-1.01	0
France	-0.79	0	-1.84	0
Germany	-1.19	0	-1.35	0
Italy	-1.74*	0	-2.08	0
Japan	-1.48	0	-1.67	0
Netherlands	-0.32	1	-1.36	1
Norway	0.56	1	-0.77	1
Spain	-0.76	0	-1.60	0
Sweden	-1.66*	0	-2.31	0
Switzerland	-0.74	2	-0.81	2
UK	-0.05	0	-1.51	0
US	-1.11	0	-1.21	0

Notes: Data from 1975:1 to 2002:3. The lag length is selected by Modified SIC, with maximal lag length equal to 6. Asterisks denotes significance at the *10%, ** 5% , and *** 1% levels. The critical values corresponding to these significance levels are (-1.61, -1.94, -2.59) for the demeaned seriesand (-2.73, -3.02, -3.57) for the detrended series, respectively.

Table 9: ADF test: RER based on normalized unit labor cost

	Demeaned		Detrended		OLS
Country	ADF	Lags	ADF	Lags	Trend
Austria	-0.55	0	-2.79	1	-0.0002***
Belgium	-2.10	0	-1.37	0	0.0000
Canada	-2.00	1	-2.07	1	0.0000
Denmark	-2.34	0	-2.56	0	0.0000
Finland	-1.30	0	-4.09	0	-0.0004***
France	-0.38	0	-3.44*	0	-0.0002***
Germany	-1.22	0	-2.14	0	0.0001*
Italy	-2.45	0	-2.39	0	0.0000
Japan	-1.91	0	-2.39	0	0.0001
Netherlands	-1.75	0	-1.20	0	0.0000
Norway	-1.15	1	0.01	0	0.0001
Spain	-2.48	0	-2.44	0	0.0000
Sweden	-1.15	0	-3.01	0	-0.0002***
Switzerland	-0.43	2	-2.51	2	0.0002***
UK	-1.18	0	-1.64	0	0.0000
US	-1.25	0	-1.57	0	-0.0001

Notes: Data from 1975:1 to 2002:3. The lag length is selected by Modified SIC, with maximal lag length equal to 6. Asterisks denotes significance at the *10%, ** 5% , and *** 1% levels. The critical values corresponding to these significance levels are (-2.58, -2.89, -3.49) for the demeaned series and (-3.15, -3.45, -4.04) for the detrended series, respectively.

Table 10: DF-GLS test: RER based on normalized unit labor cost

	Demeaned		Detrended	
Country	DF-GLS	Lags	DF-GLS	Lags
Austria	-0.53	0	-1.38	1
Belgium	0.71	1	-0.87	0
Canada	-1.80*	1	-1.90	1
Denmark	-2.32*	0	-2.41	0
Finland	-0.83	0	-0.98	0
France	-0.42	0	-1.55	0
Germany	-1.14	0	-1.32	0
Italy	-1.56	0	-1.96	0
Japan	-1.50	0	-1.70	0
Netherlands	0.07	0	-1.21	0
Norway	1.30	1	-0.38	0
Spain	-0.83	0	-1.91	0
Sweden	-1.56	0	-2.40	0
Switzerland	-0.68	2	-0.88	2
UK	0.07	0	-1.61	0
US	-1.02	0	-1.15	0

Notes: Data from 1975:1 to 2002:3. The lag length is selected by Modified SIC, with maximal lag length equal to 6. Asterisks denotes significance at the *10%, ** 5% , and *** 1% levels. The critical values corresponding to these significance levels are (-1.61, -1.94, -2.59) for the demeaned seriesand (-2.73, -3.02, -3.57) for the detrended series, respectively.

Table 11: ADF test: RER based on value added

	Demeaned		Detrended		OLS
Country	ADF	Lags	ADF	Lags	Trend
Austria	-1.42	4	-2.30	3	-0.0002**
Belgium	-1.73	0	-1.04	0	0.0000
Canada	-1.65	1	-1.69	0	0.0000
Denmark	-0.35	0	-1.62	0	0.0001*
Finland	-0.92	0	-1.57	0	-0.0001
France	-2.05	0	-2.14	0	0.0000
Germany	-2.06	6	-2.02	3	0.0001
Italy	-1.82	0	-1.81	0	0.0000
Japan	-2.06	1	-1.65	0	0.0001
Netherlands	-2.09	0	-1.76	0	0.0000
Norway	-1.45	0	-2.28	0	0.0001
Spain	-2.21	0	-2.27	0	0.0001
Sweden	-2.33	0	-2.27	0	0.0000
Switzerland	-2.92	0	-2.94	0	0.0000
UK	-1.75	0	-1.77	0	0.0000
US	-1.08	0	-1.57	0	-0.0001

Notes: Data from 1975:1 to 1997:1. The lag length is selected by Modified SIC, with maximal lag length equal to 6. Asterisks denotes significance at the *10%, ** 5% , and *** 1% levels. The critical values corresponding to these significance levels are (-2.59, -2.90, -3.51) for the demeaned series and (-3.16, -3.46, -4.07) for the detrended series, respectively.

Table 12: DF-GLS test: RER based on value added

	Demeaned		Detrended	
Country	DF-GLS	Lags	DF-GLS	Lags
Austria	-0.99	4	-1.90	3
Belgium	-0.85	4	-0.83	0
Canada	-1.46	0	-1.55	0
Denmark	-0.27	0	-1.19	0
Finland	-0.69	0	-1.54	0
France	-2.00**	0	-2.17	0
Germany	-1.84*	6	-2.05	3
Italy	-1.52	0	-1.78	0
Japan	-1.20	1	-1.80	0
Netherlands	-0.53	0	-1.33	0
Norway	-1.57	0	-2.26	0
Spain	-1.63*	0	-2.31	0
Sweden	-1.65*	0	-2.08	0
Switzerland	-2.92***	0	-3.00*	0
UK	-0.60	0	-1.57	0
US	-0.84	0	-1.50	0

Notes: Data from 1975:1 to 1997:1. The lag length is selected by Modified SIC, with maximal lag length equal to 6. Asterisks denotes significance at the *10%, ** 5% , and *** 1% levels. The critical values corresponding to these significance levels are (-1.61, -1.94, -2.59) for the demeaned seriesand (-2.79, -3.08, -3.64) for the detrended series, respectively.

Table 13: ADF test: RER based on CPI over WPI

	Demeaned		Detrended		OLS
Country	ADF	Lags	ADF	Lags	Trend
Austria	-0.01	1	-1.47	0	0.0001
Belgium	-0.80	1	-1.59	1	0.0001*
Canada	-1.49	1	-1.38	1	0.0000
Denmark	-1.72	0	-1.88	0	0.0000
Finland	-1.42	0	-1.61	0	0.0000
France	-0.71	0	-1.70	0	0.0001
Germany	-0.96	1	-1.51	0	0.0001*
Italy	-1.10	1	-1.92	0	0.0002*
Japan	-1.38	1	-1.81	0	0.0002
Netherlands	-1.00	1	-1.25	0	0.0001
Norway	-1.16	1	-1.33	0	0.0001
Spain	-1.04	1	-1.32	0	0.0001
Sweden	-1.12	1	-1.25	0	0.0001
Switzerland	-0.75	1	-1.20	0	0.0001
UK	-0.82	1	-1.25	0	0.0001
US	-0.92	0	-1.38	0	0.0001

Notes: Data from 1975:1 to 1998:4. The lag length is selected by Modified SIC, with maximal lag length equal to 6. Asterisks denotes significance at the *10%, ** 5% , and *** 1% levels. The critical values corresponding to these significance levels are (-2.58, -2.89, -3.50) for the demeaned series and (-3.15, -3.46, -4.06) for the detrended series, respectively.

Table 14: DF-GLS test: RER based on CPI over WPI

	Demeaned		Detrended	
Country	DF-GLS	Lags	DF-GLS	Lags
Austria	0.69	2	-1.36	0
Belgium	-0.66	1	-1.53	1
Canada	-1.51	1	-1.50	1
Denmark	-1.55	0	-1.64	0
Finland	-0.35	0	-1.60	0
France	-0.43	0	-1.57	0
Germany	-0.91	1	-1.86	1
Italy	-0.59	1	-1.81	0
Japan	-0.16	1	-1.84	0
Netherlands	-0.70	1	-1.30	0
Norway	-0.87	1	-1.36	0
Spain	-0.79	1	-1.34	0
Sweden	-0.92	1	-1.27	0
Switzerland	-0.48	1	-1.21	0
UK	-0.52	1	-1.27	0
US	-0.93	0	-1.41	0

Notes: Data from 1975:1 to 1998:4. The lag length is selected by Modified SIC, with maximal lag length equal to 6. Asterisks denotes significance at the *10%, ** 5% , and *** 1% levels. The critical values corresponding to these significance levels are (-1.61, -1.94, -2.59) for the demeaned seriesand (-2.79, -3.08, -3.64) for the detrended series, respectively.

Table 15: Countries with PPP evidence

	ADF		DF-GLS	
RER Proxy	Demeaned	Detrended	Demeaned	Detrended
Exp.unit v.	France Sweden	Switzerland	France Germany Italy Netherland	France Switzerland
WPI	-	-	Finland France Germany Italy Spain Switzerland	-
CPI	-	-	Denmark Finland Italy Norway	Switzerland
Labor u.c.	-	France Sweden	Denmark Italy Sweden	-
Norm.l.u.c.	-	France	Canada Denmark	-
Value add.	Switzerland	-	France Germany Spain Sweden Switzerland	Switzerland
CPI/WPI	-	-	-	-