r b e

A Note on Purchasing Power Parity and the Choice of Price Index

Cristina Terra*, Ana Lucia Vahia

Contents: 1. Introduction; 2. Data and methodology; 3. Empirical Results; 4. Concluding Remarks; A. Tables.

Segundo o argumento de Paridade do Poder de Compra, o poder de compra das moedas deveria ser o mesmo entre as economias para uma mesma cesta de bens comerciáveis. Neste artigo promovemos uma competição entre seis diferentes indíces de preços para investigar qual deles apresenta maior evidência de PPC, melhor satisfazendo, portanto, esse critério. Utilizamos o Valor Unitário das Exportações, o Índice de Preços por Atacado, o Custo Unitário do Trabalho, o Custo Unitário do Trabalho Normalizado e o Índice de Preços ao Consumidor para 16 países entre 1975 e 2002. Em testes de raiz unitária, o Índice de Preços por Atacado foi o índice para o qual a PPC foi encontrada para o maior número de países. Não foi encontrada nenhuma evidência de PPC para a razão entre o Índice de Preços ao Consumidor e o por Atacado.

The PPP argument states that the currencies purchasing power should be the same across countries for the same basket of tradable goods. We run a horse race among six different price indices to see which one yields higher PPP evidence, and, therefore, better fits this criterion. Using export unit values, WPIs, value added deflators, unit labor costs, normalized unit labor costs and CPIs for 16 countries from 1975 to 2002, unit root tests show that WPI was the index for which PPP evidence was found for the larger number of countries. No evidence of PPP was found for the ratio CPI/WPI.

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. There is an extensive 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 real exchange rate (RER). For surveys on the empirical literature on PPP,

see Froot and Rogoff (1995), Rogoff (1996), Sarno and Taylor (2002), and Taylor and Taylor (2004). 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 fit these two features. Hence, we run a horse race among six different price indices available from the IMF database to see which one would yield higher PPP evidence. We use 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 is tested using both the ADF and the DF-GLS unit root test of the RER series, for a sample of 16 industrial countries, with quarterly data from 1975 to 2002.

The RER measured as WPI ratios is the one for which PPP evidence is found for the larger number of countries: six out of sixteen when we use DF-GLS test with demeaned series. This is an indication that, from all indices used, WPI seems to be the one with larger composition of tradable goods and with one least variation in its basket of goods composition across countries. The second best RER measure was the value added deflators. On the one hand, this is an index that includes the cost of all factors of production. On the other hand, the index composition may vary substantially across countries due to the lack of cross-country comparability.

There are studies that test the PPP hypothesis for different price indices such as Dornbush (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 (2000) 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.

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 tradadable goods with similar 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, deterministic trends were found to be significant, possibly indicating some Balassa-Samuelson effect. Fourth, for the RER measured as the ratio of foreign CPI and domestic WPI, we find no evidence of PPP holding. This is consistent with the idea that CPI has a large share of nontradable goods which are not arbitraged across countries.

This paper is organized as follows. The data and the methodology are presented in section two. Section three presents the empirical results, and section four concludes.

2. DATA AND METHODOLOGY

We use the multilateral real exchange rate to PPP testing to take into account the behavior of all the relevant bilateral exchange rates. Following the IMF methodology, the RER was computed as in the following equation:

$$RER_i = \Pi_{j \neq i} \left[\frac{E_i P_i}{E_j P_j} \right]^{W_{ij}} \tag{1}$$

where E_i is the nominal exchange rate, period-average US dollars per unit of national currency, P_i is the price index and W_{ij} is the weight attached by country i to country j.[1]

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 have quarterly data for 16 industrialized countries:

---

[1]The description of the methodologies for computation of these weights for each price index is described in the working paper version of this article, Terra and Vahia (2005).

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. The WPI and the value added deflator range from 1975 to 1997, and the 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 that changes with export growth of new entrants with low prices. As the export pattern differs substantially across countries, so does the composition of the export unit value.

The consumer price indices have 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, CPIs and WPIs include 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.

The 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 advantage 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 costs, 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.

As for the methodology, we need to test the nonstationary of the RER series. When the RER is nonstationary, the series will present a unit root and the PPP hypothesis is rejected. Evidence against unit root behavior emerges when the RER fluctuates around a constant mean, 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.

We use both the traditional augmented Dickey-Fuller test (ADF) and the generalized-least-square version of the Dickey Fuller test (DF-GLS) to test for the nonstationarity of the RER. We believe that the DF-GLS is the most suitable test for, basically, two reasons. First, it is a solution suggested by the

literature for the power problem[2] (Taylor, 2002). According to Elliot et al. (1996), "the Dickey-Fuller t-test applied to a locally demeaned 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." Second, it allows for deterministic trends, in the spirit of the Balassa-Samuelson effect.

3. 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 ADF and the DF-GLS tests. The results are summarized in Tables 1 and 2. Table 1 presents the list of countries for which the unit root null hypothesis can be rejected for each of the price indices used and for the two tests performed.[3] In Table 2, the results of a simple OLS regression on a constant and a trend is presented for all countries and price indices. These results are used to verify the existence of a deterministic trend in the series.

We start with PPP testing for the RER based on export unit values. The results of the ADF unit root tests are in the first two columns of the first box of 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 (on the first column of Table 2) indicates the presence of a trend for Canada, Spain 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 estimation without trend, that is, in the 'demeaned' result.

Differently from the ADF test, the DF-GLS test rejects the unit root null for Sweden (last two columns of Table 1). 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 value 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 the second box of Table 1. 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 the larger set of countries.

As for the RER series constructed as CPI ratios (in the third box of Table 1), 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

---

[2]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 at a very slow speed of convergence, that is, the value of γ is very near zero.

[3]Terra and Vahia (2005) present the full results for all tests.

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 are more heavily weighted with nontradable goods than tradable goods, when compared with WPIs. 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 the fourth and fifth box of Table 1. Adding a trend to the estimation results in the 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 1 shows that the unit root null is rejected for Denmark, Italy and Sweden, for the RER based on unit labor cost. For the normalized series, unit root is rejected only for Canada. We cannot reject unit roots for any of the detrended estimations, for the two sets of RER series. This means that no deterministic trend explain the unit root evidence.

These results indicate 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 the second to last box of Table 1, detect no unit root. The DF-GLS, on the other hand, rejects the unit root null for five countries: France, Germany, Spain, Sweden and Switzerland. These results 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, as shown in the last box of Table 1. This proxy for RER suffers from two of the problems that could causes PPP deviations: 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).

Note that, for all series studied, 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 when we use DF-GLS test with demeaned series. 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 constucted as the ratio between foreign countries CPIs and domestic country WPI.

Looking from the countries' perspective, France is the country for which PPP evidence was found for 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.

4. CONCLUDING REMARKS

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 similar basket of goods for all countries. Second, when using export unit values, the PPP is verified for only four countries. This index includes only goods that are actually traded by the country, hence its basket of goods composition most probably differs across countries to a greater extent, compared to the other indices. Third, deterministic trends were found to be significant, possibly indicating some Balassa-Samuelson effect. Fourth, for the RER measured as the ratio of foreign CPI and domestic WPI, we find no evidence of PPP holding. This is consistent with the idea that CPI has a large share of nontradable goods which are not arbitraged across countries.

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

Bibliography

Chinn, M. D. (2000). Before the fall: were East Asian currencies overvalued? *Emerging Markets Review*, 1(2):101–126. Available at http://ideas.repec.org/a/eee/ememar/v1y2000i2p101-126.html.

Dornbush, R. (1987). Purchasing power parity. In John Eatwell, M. M. & Newman, P., editors, *The New Palgrave: A Dictionary of Economics*, volume 3, pages 1075–85. Stockton Press, 3 edition.

Elliot, G., Rothenberg, T. J., & Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, 64(4):813–36. Available at `http://ideas.repec.org/a/ecm/emetrp/v64y1996i4p813-36.html`.

Froot, K. A. & Rogoff, K. (1995). Perspectives on ppp and long run real exchange rates. In Grossman, G. M. & Rogoff, K., editors, *Handbook of International Economics*, volume 3. Amsterdam: North-Holland.

Rogoff, K. (1996). The purchasing power parity puzzle. *Journal of Economic Literature*, 34(2):647–668. Available at `http://ideas.repec.org/a/aea/jeclit/v34y1996i2p647-668.html`.

Sarno, L. & Taylor, M. P. (2002). Purchasing power parity and the real exchange rate. *IMF Staff Papers*, 49(1):5. Available at `http://ideas.repec.org/a/imf/imfstp/v49y2002i1p5.html`.

Taylor, A. M. (2002). A century of purchasing-power parity. *The Review of Economics and Statistics*, 84(1):139–150. Available at `http://ideas.repec.org/a/tpr/restat/v84y2002i1p139-150.html`.

Taylor, A. M. & Taylor, M. P. (2004). The purchasing power parity debate. *Journal of Economic Perspectives*, 18(4):135–158.

Terra, C. & Vahia, A. L. (2005). Purchasing power parity: The choice of price index. Ensaios Econômicos 588, EPGE.

A. TABLES

Table 1 – Countries with PPP evidence

RER Proxy	ADF		DF-GLS	
	Demeaned	Detrended	Demeaned	Detrended
Exp. unit v.	France Sweden	Switzerland	France Germany Italy Netherlands	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. v.	-	France	Canada Denmark	-
Value add.	Switzerland	-	France Germany Spain Sweden Switzerland	Switzerland
CPI/WPI	-	-	-	-

Table 2 – OLS Trend

Country	Exp. unit v.	WPI	CPI	Labor u. c.	Norm. l. u. v.	Value add.	CPI/WPI
Austria	0,0000	-0,0001**	0,0000	-0,0002***	-0,0002***	-0,0002**	0,0001
Belgium	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001*
Canada	-0,0002**	0,0000	-0,0001	0,0000	0,0000	0,0000	0,0000
Denmark	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001*	0,0000
Finland	0,0000	0,0000	0,0000	-0,0004***	-0,0004***	-0,0001	0,0000
France	0,0000	0,0000	0,0000	-0,0002***	-0,0002***	0,0000	0,0001
Germany	0,0000	0,0000	0,0000	0,0001*	0,0001*	0,0001	0,0001*
Italy	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002*
Japan	0,0001	0,0001	0,0001	0,0000	0,0001	0,0001	0,0002
Netherlands	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001
Norway	-0,0001	0,0000	0,0000	0,0000	0,0001	0,0000	0,0001
Spain	0,0001*	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001
Sweden	0,0000	0,0000	-0,0001	-0,0002***	-0,0002***	0,0000	0,0001
Switzerland	0,0004***	0,0001	0,0001*	0,0002***	0,0002***	0,0000	0,0001*
UK	0,0001	0,0001	0,0001	0,0000	0,0000	0,0000	0,0001
US	0,0000	-0,0001	0,0000	-0,0001	-0,0001	-0,0001	0,0001

Notes: Data from 1975:1 to 1998:2 to export unit value, from 1975:1 to 1997:1 to wholesale price index, from 1975:1 to 2002:3 to consumer price index, from 1975:1 to 2002:3 to unit labor cost, from 1975:1 to 2002:3 to normalized unit labor cost, from 1975:1 to 1997:1 to value added, from 1975:1 to 1998:4 to CPI over WPI. Asterisks denotes significance at the * 10%, ** 5%, and *** 1% levels.