Empir Econ (2012) 42:713–744
DOI 10.1007/s00181-010-0441-0

Digging out the PPP hypothesis: an integrated empirical coverage

Miguel de Carvalho · Paulo Júlio

Received: 26 September 2009 / Accepted: 29 September 2010 / Published online: 16 December 2010

Abstract We use several popular tests to test the validity of the Purchasing Power Parity (PPP) hypothesis. In particular, we analyze four classes of tests—standard univariate unit root tests, co-integration, panel unit root tests, and unit root tests for nonlinear frameworks—for a dataset consisting of 20 bilateral exchange rates. Through this approach, we ascertain the effectiveness of each methodology in assessing the validity of PPP. Overall, our results suggest little evidence to support PPP. Among the conducted tests, the Panel Analysis of Nonstationarity in the Idiosyncratic and Common components (PANIC) provides the richest insights by disentangling the possible sources of non-stationarity of real exchange rates. The relevance of using price indices with different characteristics is also pinpointed.

Keywords PPP · Real exchange rate · Unit roots · Co-integration · Panel · Nonlinear models · Cross-sectional dependence

1 Introduction

The Purchasing Power Parity (PPP) hypothesis is one of the most important and intuitive ideas in economics, serving as basis to set up many open economy models; yet,

M. de Carvalho
Ecole Polytechnique Fédérale de Lausanne, Institute of Mathematics, EPFL SB-IMA Station 8, 1015 Lausanne, Switzerland
e-mail: miguel.carvalho@epfl.ch

P. Júlio (✉)
Gabinete de Estratégia e Estudos, Ministry of Economy, Innovation and Development, Av. da República, 79, 1.º, 1050–246 Lisbon, Portugal
e-mail: paulo.julio@gee.min-economia.pt

it keeps challenging the economic community worldwide. Intuitively, PPP is a simple relationship which states that the price levels in two different countries should be the same, when converted to a common currency. Otherwise large profit opportunities from cross-border trade would arise, and economic agents would engage in arbitrage activities leading to corrections either in prices or nominal exchange rates, or both, until price equalization is attained. To put this differently, while short-run deviations are admissible in light of PPP, the Real Exchange Rate (RER) should be stationary, such that the purchasing power of two different currencies does not deviate permanently from its long-run equilibrium.

Despite this simple and intuitive idea, which economists undoubtedly expect to hold, data often refute the idea that deviations from the equilibrium RER are a simple temporary phenomenon. This apparent inconsistency between theory and evidence has raised an extensive discussion, feeding many studies and supplying the literature with several statistical tests and methodologies (Rogoff 1996; Taylor 2001; Sarno and Taylor 2002). From simple univariate unit root tests to long-span data studies, from co-integration to panel tests, PPP has been submitted to almost all types of analysis and scrutiny, and still evidence seems to be mixed at best and characterized by fairly significant levels of persistence, compared to what the theoretical arbitrage relationship would suggest.

More recently, nonlinear methods have deepened the discussion, providing stronger evidence in favor of PPP. Intuitively, nonlinear methods are able to capture different speeds of adjustment, depending on the misalignment presented by the RER. This approach brings into the empirical analysis frictions in international markets, such as trade barriers and transport costs, which introduce noise in arbitrage conditions. Stated differently, nonlinear methods allow us to postulate that the RER follows a random walk for small deviations from equilibrium, but becomes increasingly mean-reverting the further away it is from its long-run equilibrium. This observation is in line with the fact that larger profit opportunities from cross-border trade only arise when the deviation from economic fundamentals is significant.

In this article, we provide an empirical coverage of the main and most relevant empirical methods presented in the literature on PPP. Instead of focusing on one method or on some minor modifications to existing methods, we compare the results of several well-explored and well-understood methodologies, identifying the gains and the flaws of each one and assessing their successes and failures in testing the validity of PPP. At the same time, we connect our results to the major discoveries made so far in the field. Hence, this article is simultaneously a concise survey of the literature and an empirical applied paper, where the most prominent empirical methods are first presented and discussed, then applied and integrated within the literature and afterwards critically evaluated. We also emphasize the relevance of using distinct price indices, namely the Consumer Price Index (CPI) and the Producer Price Index (PPI), on the different tests we undertake, since the number of tradable goods in the latter is larger than in the former.[1]

[1] For instance, Coakley et al. (2005), using panel unit root tests, found that PPP is mostly supported when PPI is used, but mostly rejected when prices are based on CPI. In another study, Fleissig and Strauss (2000) use panel unit root tests for six distinct price indices, and conclude that PPP is generally supported, although

Our analysis relies on four classes of tests—standard univariate unit root tests, co-integration, panel unit root tests and unit root tests for nonlinear frameworks—for a dataset consisting of 20 bilateral exchange rates. Overall, our results suggest little evidence supporting the PPP hypothesis, both for CPI and PPI. The richest insight on the PPP hypothesis is offered by the Panel Analysis of Nonstationarity in the Idiosyncratic and Common components (PANIC). This method enables us to decompose any series into a pervasive component, common to all series, and an idiosyncratic component, which is series specific, and to test for the presence of unit roots in each of these components separately. Through the application of this method, we conclude that both aggregate shocks and individual specific factors contribute to deviating the RER away from its long-run equilibrium value. Another method which also provides interesting insights is the KSS test, a unit root test specifically designed for the Smooth Transition Autoregressive (STAR) model. Although many authors suggest that non-linear adjustments support PPP, our results provide little evidence in favor of a stable and globally nonlinear RER for the majority of countries in our dataset.

This article is structured as follows. Section 2 presents the theoretical foundations of PPP and the data used in the empirical tests. In Sect. 3 we select a set of univariate unit root tests to analyze the PPP hypothesis. Co-integration methods are described and employed in Sect. 4. In Sect. 5, the PPP hypothesis is dissected from the stance of panel unit root tests. The use of unit root methods for nonlinear models is outlined and applied in Sect. 6. Section 7 concludes.

1.1 Literature review

In the past few years, a vast empirical literature on PPP has emerged, taking the advantage of recently developed econometric techniques. Surveys on the empirics of PPP include Froot and Rogoff (1995), Rogoff (1996), Sarno and Taylor (2002), Taylor and Taylor (2004), and Taylor (2003, 2006). Some early studies, which amount testing for the presence of a unit root in RERs through univariate unit root tests, include, among others, Roll (1979), Edison (1985), and Meese and Rogoff (1988). In general, this methodology did not provide significant support in favor of a stable long-run RER.

The research community, however, has cast many doubts both on the validity and adequacy of these tests. First, since data on prices and nominal exchange rates are usually integrated of order one, the suitability of many univariate unit root tests seems inferior to co-integration. This allows us to test if there exists a linear combination of two or more non-stationary series which is itself stationary, providing superior adequacy as compared to traditional univariate unit root tests. Examples of this approach include, among others, Corbae and Ouliaris (1988), Taylor (1988), Mark (1990), Fisher and Park (1991), Cheung and Lai (1993), Chen (1995), and Culver and Papell (1999). In general, these studies were able to find some support for a stable long-run relationship between nominal exchange rates and prices, but did not find significant evidence

Footnote 1 continued
the speeds of adjustment differ considerably across different price indices. Along different lines, Jenkins and Snaith (2005) use consumer price sub-index data in a multivariate co-integration framework, and conclude that the failure of PPP can be attributed to the inclusion of non-traded goods in the overall index.

that the theoretical co-integration vector is included in the co-integrating space, which can be interpreted as a violation of PPP.

Second, if the true process that drives RERs is endowed with a “near unit root,” such that shocks have an effect over several years, as some evidence suggests (Rogoff 1996), then univariate tests might induce a large type II error, and many more observations may be necessary to dismiss the presence of unit roots. Two main different alternatives have been proposed in the literature to deal with this “power problem”—long-span data studies and panel unit root tests.

The use of long-span data is apparently the easiest way to increase the power of unit root tests. For instance, Frankel (1986), Edison (1987), Lothian and Taylor (1996), Taylor (2002), and Sarno and Valente (2006) used data for over 60 years, and had some remarkable successes when confronting the null hypothesis of non-stationary RERs with the alternative of stationarity. However, Wallace and Shelley (2006) find only mixed evidence in favor of PPP. Engel (2000) questions the validity of these tests, arguing that they present serious size bias which may lead to wrong conclusions. A different source of bias arises when the same model is assumed to be valid under distinct exchange rate regimes. It is well-known that the behavior of RERs is highly dependent on whether the exchange rate is allowed to float or not, and traditional unit root tests using long-span data did not take into account the structural shifts created by the different regimes in place over the past century. Finally, this approach can only be applied to countries where such long data series are available, which greatly restricts the analysis and cross-country comparisons.

An alternative way to increase the size and power of unit root tests is to expand the cross-section dimension of the database, by gathering several countries in a panel of observations. Along these lines, Wu (1995), Frankel and Rose (1996), Oh (1996), and Papell (1997), among others, were able to provide some evidence favoring mean-reversion of RERs, although at very slow rates. Sarno and Taylor (1998) and Taylor and Sarno (1998), however, argue that some of these studies incorrectly interpret the null hypothesis of non-stationarity. In particular, they note that the rejection of joint non-stationarity may occur even if only one of the series in the panel is stationary. This may have lead to some wrong conclusions in early panel studies. Moreover, the results of these tests are strongly dependent on the numeraire currency. Conversely, O’Connell (1998) and Smith et al. (2004), for instance, conclude that no evidence favoring PPP can be found, after correcting the tests for cross-sectional dependence. Koedijk et al. (1998), Fleissig and Strauss (2000), and Kuo and Mikkola (2001) also consider cross-sectional dependence in their analysis, but, in general terms, are able to support mean-reversion in RERs, although the rates of convergence differ significantly across countries.

A related and more recent approach is the multivariate co-integration test, which applies co-integration techniques to panel data in order to exploit the advantages of each of these methodologies. Examples of applications of these techniques include Jacobson and Nessén (2004) and Jacobson et al. (2008), who have concluded that prices and exchange rates appear to be co-integrated, but with a co-integrating vector different from what the theory suggests, a conclusion that is in line with the findings of standard co-integration tests. These results hint that measurement errors in prices and transaction costs may create a wedge between prices and nominal exchange rates, such that only a weaker version of PPP is supported by the data.

At this stage, and after an exhaustive application of panel and co-integration tests, no clear consensus has emerged in the research community; in fact, the debate about the validity of PPP seems to be more vigorous than ever, with contradictory evidence continually flooding the literature. Even when supportive evidence for a stable RER could be found, the rate of convergence to the long-run equilibrium was usually surprisingly low (Rogoff 1996). This high degree of persistence in RERs has only been considered as satisfactorily addressed when some authors (e.g., Goldberg et al. 1997; Michael et al. 1997; Sarantis 1999; Taylor and Peel 2000; Taylor et al. 2001; Sollis et al. 2002; Liew et al. 2003; Liew 2004) started to bring nonlinear adjustments to PPP. In general, all these studies show that RERs are characterized by strong nonlinearities, and conclude that economic forces driving mean-reversion are stronger the farther the RER is from its long-run equilibrium value. The adequacy of nonlinear methods seems superior to that of traditional methodologies, which rely on a uniform autoregressive parameter that does not incorporate different corrective pressures that characterize the transition to equilibrium.

2 Preliminaries

2.1 The purchasing power parity and the law of one price

The main building block behind PPP is the Law of One Price (LOP). The LOP, in its absolute version, postulates that the same good should have the same price across countries, if prices are expressed in a common currency. More formally

$$S_t \cdot P_{i,t}^* = P_{i,t};\ i = 1, 2, \ldots, N,\ \forall t \tag{1}$$

where $P_{i,t}$ is the price of good i at time t in the domestic country, $P_{i,t}^*$ is the price of the same good at time t in the foreign country, and S_t denotes the nominal exchange rate, defined as the amount of domestic currency needed to buy one unit of foreign currency.

The relative version of the LOP postulates a weaker condition, that the relative price of good i is constant through time when expressed in a common currency. Formally

$$\frac{S_t \cdot P_{i,t}^*}{P_{i,t}} = C;\ i = 1, 2, \ldots, N,\ \forall t \tag{2}$$

where C is some positive constant. Obviously, the absolute LOP implies its relative version, but the converse is false.

The main idea behind the absolute LOP is frictionless goods arbitrage (Sarno and Taylor 2002). If goods are seen by consumers as perfect substitutes, then, in the absence of trade barriers and assuming negligible transport costs, price differences across countries originate profit opportunities, which will drive price adjustments until price equalization is attained. The relative LOP weakens these assumptions, by allowing prices in the foreign country to differ from the prices in the home country by $(C-1)\cdot 100\%$.

Using Eqs. 1 and 2, both the absolute and the relative versions of PPP can be obtained immediately. Suppose we build a price index in the following way[2]

$$P_t = \sum_{i=1}^{N} \alpha_i P_{i,t} \tag{3}$$

where $\alpha_i \in [0, 1]$, with $\sum_{i=1}^{N} \alpha_i = 1$, and N represents the number of tradable goods satisfying the LOP. If the weights α_i are the same for the domestic and foreign countries, then absolute PPP follows by multiplying both sides of Eq. 1 by α_i and summing up, which yields

$$S_t \cdot P_t^* = P_t \tag{4}$$

or, taking the logarithm on both sides

$$s_t + p_t^* - p_t = 0 \tag{5}$$

where $s_t = \log(S_t)$, $p_t^* = \log(P_t^*)$, and $p_t = \log(P_t)$.

Equation 5 simply states that the purchasing power of one unit of currency is the same in both countries, when converted to the same monetary unit. To put this differently, if absolute PPP holds, then any economic agent can buy the same basket of goods with the same amount of money in both countries. This relationship is supported by stronger assumptions than the ones invoked by the LOP; namely the weights used to compute the price index must be the same in both countries. This is a much stricter condition, which is unlikely to hold in practice, not only due to distinct statistical methodologies used, but also because different sets of goods may be used by countries in the construction of such measure.

Applying the same mechanism to Eq. 2, relative PPP also follows immediately

$$s_t + p_t^* - p_t = c \tag{6}$$

where $c = \log(C)$. Condition (6) presents two main attractions as compared to (5). Firstly, it is based on the relative LOP, which imposes weaker assumptions by admitting price differences across countries. Secondly, it is much easier to test empirically than its absolute version, since the data collected on prices is based on indices rather than on levels, which creates a wedge between the relative prices of different countries that can only be captured through the parameter c. It is therefore the relationship provided by (6) that will be the object of our study.

[2] Conversely, $P_t^* = \sum_{i=1}^{N} \alpha_i P_{i,t}^*$.

2.2 Stationarity or random walk! why is it so important?

Although PPP as stated in Eq. 6 is an identity, it is hard to believe that it holds continually every period, as the volatility of nominal exchange rates is much larger then the volatility presented by prices. However, most economists believe that PPP constitutes an anchor for RERs in the long-run, such that any shock to the PPP relationship in (6) eventually dies out. If we let the logarithm of the RER be

$$e_t = s_t + p_t^* - p_t \tag{7}$$

this amounts to say that e_t must display the property of mean-reversion, such that deviations from the long-run equilibrium are nothing more than a temporary phenomenon with no long-run repercussions. Hence, testing PPP reduces to test the stationarity of the RER.

A natural question is why there has been so much concern about this issue. Many important reasons justify such extensive research on PPP. For instance, Sarno and Taylor (2002) and Rogoff (1996) point out that, if shocks are highly persistent, so that the RER is close to a random walk, then they must be originated from the real side (e.g., technology shocks), while, if shocks show little persistence, then they must be originated by aggregate demand (e.g., monetary policy). Also, much open economy macroeconomics is based on PPP, and its invalidity implies that all research based upon it would have to be reassessed (Taylor 1995; Sarno 2005). Moreover, PPP is used not only to determine the degree of misalignment of nominal exchange rates, but also to compare national income levels (Sarno and Taylor 2001), and this only makes sense if the relative purchasing power of two currencies does not change over time.

2.3 Data

The data we used were gathered from the International Monetary Fund's *International Financial Statistics* (on-line) database. Both quarterly data on bilateral exchange rates of the national currency against the U.S. dollar and on two price measures—the CPI and the PPI—were collected, for the period 1973:1–2007:4. The base year for both price indices is 2000. The analysis comprises 20 developed countries: Canada, Australia, Japan, New Zealand, United Kingdom, Switzerland, Sweden, Norway, Denmark, Austria, Belgium, Finland, France, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, and Spain. (The analysis concerning PPI excludes France, Greece, and Portugal, for which data were only available for a short period of time. For Belgium, Italy, Luxembourg, Ireland, and Norway, some initial observations are missing, but the analysis was still conducted for the available time span.) Whenever necessary, the Euro–Dollar exchange rate was converted to the exchange rate of the former currency of the country against the Dollar.[3] All variables were put into natural logarithms before the analysis. In our notation, p_t^* is the price index in the U.S., p_t

[3] Through a visual inspection of the series, we noted that the creation of the Euro-Zone has brought no change to the long-run behavior of RERs. Structural breaks are also absent from the series. Moreover, in order to be confident that our results were not biased by this approach, we also undertook several of the

is the price index of the country being tested, and s_t is the amount of currency of the country being tested needed to buy 1 U.S. Dollar.

3 Standard univariate unit root/stationarity tests

In this section, we employ three simple univariate tests. We start by introducing the ADF test, a univariate unit root test whose main interest lies in motivating the subsequent analysis. Next, we introduce the DF-GLS, a modified version of the Dickey–Fuller test which allows a considerable increase in power. Finally, we present the results for the KPSS test, a stationarity test that assumes a contrary specification for the null and alternative hypothesis. Our analysis here allows us to present the main problems underlying the PPP hypothesis as pointed out in the literature, serving as a motivation for more complex frameworks that need to be addressed. Unless stated otherwise, we use a significance level of 5%.

3.1 Theoretical background

3.1.1 The ADF test

The ADF test evaluates the null hypothesis of a unit root against the alternative that the process is stationary. It presupposes the following specification of the data generating process

$$\Delta e_t = \eta_0 + \alpha e_{t-1} + \sum_{i=1}^{p} \eta_i \Delta e_{t-i} + u_t \tag{8}$$

where u_t is an i.i.d. error term. The ADF statistic provides us the means to test the null of $\alpha = 0$ against the alternative that $\alpha < 0$ in this auxiliary regression.

3.1.2 The DF-GLS test

One of the main hindrances of several univariate unit root tests is their low power. Elliott et al. (1996) (hereinafter ERS) therefore proposed a modified Dickey–Fuller test—the DF-GLS test—that can be performed by testing $\alpha = 0$, against the alternative $\alpha < 0$, in the regression

$$\Delta e_t^d = \alpha e_{t-1}^d + \sum_{i=1}^{p} \eta_i \Delta e_{t-i}^d + u_t \tag{9}$$

where e_t^d denotes locally demeaned data. An application of this method to PPP can be found in Cheung and Lai (1998).

Footnote 3 continued
tests presented herein for the period 1973:1–1998:4. Since conclusions remained broadly unchanged, we decided to use the longest time span available in order to take advantage of more powerful tests.

3.1.3 The KPSS test

As a means to investigate the PPP hypothesis from a different perspective, we now outline a test which has its roots in the study of Kwiatkowski et al. (1992) (henceforth KPSS). In this test, stationarity is the null, rather than the alternative hypothesis. The test starts by estimating the equation

$$e_t = \mu + u_t \tag{10}$$

where u_t is an i.i.d. error term. From the residuals of this regression, $\widehat{u}_t$, we estimate the unconditional error variance, $\widehat{\sigma}_u^2$, and compute the partial sum of residuals S_t. The test statistic, based on Nabeya and Tanaka (1988), is given by $\text{LM} = \widehat{\sigma}_u^{-2} \sum_{t=1}^{T} S_t^2$. Under a non-i.i.d. setup, a consistent estimator of the long-run variance should be used instead of $\widehat{\sigma}_u^2$.

3.2 Results

Table 1 provides the results for these three univariate tests. Using the CPI (PPI), we observe that in 3 (4) out of 20 (17) cases the ADF test rejects the null of a unit root. This constitutes a very low percentage of rejections. One of the reasons given in the literature is the high persistence of RERs, whose estimated half-lives are about 3–5 years. If these estimates are correct, then RERs follow a "near unit root" process, and the power of the ADF test may be too low to reject the null hypothesis of a random walk. In order to circumvent this problem, the DF-GLS t-statistic is also presented. The results under this test improve substantially, with about 60% of rejections for CPI and 65% for PPI. Notwithstanding, they still do not discard the random walk hypothesis for a relatively large set of countries. For Belgium, Luxembourg, New Zealand, and Norway, the DF-GLS test corroborates PPP when one uses CPI, but not for PPI. These results suggest that although PPI is composed of more tradable goods, it does not necessarily improve the case for PPP for all the pairs being tested. In fact, the number of rejections is marginally identical for both price indices.

Finally, the null of stationarity can be rejected by the KPSS test in five cases for CPI and in two cases for PPI, in line with the findings in the literature (e.g., Culver and Papell 1999). In order to obtain a rich insight on the PPP hypothesis, we can compare the outcome of the DF-GLS with that of the KPSS. For CPI, these two tests yield the conclusion of stationarity for: Austria, Belgium, Denmark, Finland, France, Greece, Italy, Luxembourg, Netherlands, New Zealand, Norway; and the same conclusion of a unit root for Australia, Canada, Japan and United Kingdom. Contrary conclusions are true for Ireland, Portugal, Spain, Sweden, and Switzerland. For PPI, we conclude that the RER is stationary for Austria, Denmark, Finland, Ireland, Japan, Netherlands, Spain, Sweden, and Switzerland, and a random walk for New Zealand. For Australia, Belgium, Canada, Luxembourg, Norway, and United Kingdom the KPSS and the DF-GLS tests present contradicting evidence.

Table 1 Results of univariate unit root/stationarity tests

Country		CPI			PPI		
		ADF	DF-GLS	KPSS	ADF	DF-GLS	KPSS
Australia		−1.77	−1.36	0.67^b	−2.47	-1.82^a	0.17
Austria		-2.74^a	-2.06^b	0.14	-2.92^b	-2.76^c	0.08
Belgium		-2.71^a	-2.73^c	0.15	−2.47	-1.94^a	0.16
Canada		−2.01	−1.48	0.65^b	−2.09	-1.78^a	0.13
Denmark		-2.68^a	-2.38^b	0.10	−2.23	-2.09^b	0.19
Finland		-2.60^a	-2.49^b	0.31	-2.86^a	-2.52^b	0.37^a
France		-2.82^a	-2.83^c	0.13	n.a.	n.a.	n.a.
Greece		−2.26	-2.00^b	0.26	n.a.	n.a.	n.a.
Ireland		−2.18	-1.70^a	0.32	-3.30^b	-3.22^c	0.10
Italy		-2.76^a	-2.79^c	0.11	−2.13	-1.99^b	0.22
Japan		−2.48	−1.31	0.59^b	−2.14	-1.95^b	0.39^a
Luxembourg		-2.64^a	-2.57^b	0.17	−1.69	−1.54	0.13
Netherlands		-3.20^b	-2.95^c	0.19	-2.63^a	-2.65^c	0.22
New Zealand		-2.89^b	-2.79^c	0.08	−1.49	−1.46	0.70^b
Norway		−2.15	-2.04^b	0.12	−1.35	−1.27	0.10
Portugal		−1.85	-1.66^a	0.40^a	n.a.	n.a.	n.a.
Spain		−2.28	-1.64^a	0.17	-2.67^a	-2.62^c	0.08
Sweden		−2.32	-2.01^b	0.56^b	-2.89^b	-2.90^c	0.11
Switzerland		-3.14^b	−1.44	0.23	-3.14^b	-2.79^c	0.13
United Kingdom		-2.64^a	-1.72^a	0.57^b	−2.05	-2.02^b	1.01^c
Number of rejections	*a*	11	16	6	7	14	4
	b	3	12	5	4	11	2
	c	0	5	0	0	6	1

Notes: *a*, *b* and *c* represent rejections at 10, 5, and 1% significance levels, respectively. The ADF and DF-GLS tests consider an intercept, but no trend in data. Both use the AIC as the lag length selection criteria, with an upper bound of 8 lags. The critical values for the ADF t-stat. are approximately −2.58 at 10%, −2.88 at 5%, and −3.48 at 1% significance levels, while the critical values for the DF-GLS t-stat. are approximately −1.62 at 10%, −1.94 at 5%, and −2.58 at 1% significance levels (MacKinnon 1996). The KPSS test includes an intercept, but no trend. The Bartlett kernel method was used, and bandwidth selection was made according to the Newey–West criteria. Critical values for the LM-stat.: 0.347 at 10%, 0.463 at 5%, and 0.739 at 1% significance levels (Kwiatkowski et al. 1992)

At this point, two main conclusions can be made: the random walk hypothesis cannot be conclusively refuted for about 40% of our sample; and there is no conclusive evidence that PPI provides stronger evidence in favor of PPP as compared to CPI.

3.3 Discussion

The above results constitute a challenge to any economic theorist, who might wonder how an apparently solid economic relationship build on one of the concepts most important to economists—arbitrage—does not seem to hold for the data. In fact, the question that needs to be asked at this point is how to reconcile these empirical results, widely corroborated by related literature, with economic theory which precludes the failure of such systemic relationship. The obvious answer for any economist is that the problem must lie on the empirical side, since no coherent and justifiable economic theory seems to explain the second possibility. Following this approach, three main problems inherent to univariate unit root/stationarity tests and respective solutions have been pointed out and explored in the literature:

- *Non-stationary prices and exchange rates*: The non-stationarity of any of the elements that compose the RER makes the use of standard critical values in univariate tests inappropriate (Phillips 1987). A possible solution is to use co-integration procedures.
- *The power problem*: Most univariate unit root tests have too little power to reject the null hypothesis when it is not true (Sarno and Taylor 2002). Although the power of the DF-GLS test lies near the asymptotic power envelop, panel procedures may effectively be able to attain larger improvements in power by aggregating observations across countries.
- *The linear specification*: The linear specification may not represent correctly the adjustment process faced by RERs, giving a bias towards the non-rejection of the null hypothesis (Taylor et al. 2001). A possible solution is to consider nonlinear adjustments.

At this stage, it seems that the answer relies on one or more of these alternatives. All seem rather plausible, and in the next sections we ascertain the successes and failures of each one, by applying them to our database.

4 Co-integration

Co-integration allows us to test if a linear combination of two or more non-stationary series is stationary. If this is true, then the non-stationarity of one series exactly offsets the non-stationarity of others, and the result is a stable long-run relationship between the variables. Our co-integration approach to PPP is based on Johansen's co-integration test (see, for example, Johansen 1988, 1991, 1994, 1995; Johansen and Juselius 1990). We divide the analysis into two stages. In the first, we will look for the existence of any co-integrating relationship between prices and exchange rates, without imposing any restrictions on the co-integrating vector. This can be seen as a test to a weaker version of PPP, in which no symmetry or proportionality restrictions are imposed (Cheung and Lai 1993). If we are able to reject the null hypothesis of no co-integration, we proceed to a second stage, in which the theoretical co-integration vector is imposed and tested.

4.1 Theoretical background

We start by defining the vector $X_t = (s_t, p_t^*, p_t)'$, where prices and the nominal exchange rate are assumed to be integrated of order 1. These variables are said to be co-integrated if there exist one or, at most, two co-integrating vectors $\beta_i = (1, \beta_{1i}, \beta_{2i})'$ such that

$$\beta_i' X_t = c + \eta_{it}, \quad i = 1, 2 \tag{11}$$

where η_{it} is a stationary process with $E(\eta_{it}) = 0$, and the first element of β_i is normalized to allow identification. The term η_{it} can be interpreted as the equilibrium error, and represents deviations of $\beta_i' X_t$ from the long-run equilibrium c. If the co-integrating vector is unique with $\beta = (1, 1, -1)'$, then $\beta' X_t$ is the RER and η_{it} represents deviations from theoretical PPP.

Two distinct statistics have been proposed to find the number of co-integrating relationships. In the trace test, the null hypothesis of r co-integrating relationships is tested against a general alternative of $r + k$ co-integration vectors. In the the maximum eigenvalue test, the existence of r co-integration vectors is tested against the specific alternative of $r + 1$ co-integration vectors. Finally, in order to ascertain whether the PPP vector $\beta = (1, 1, -1)'$ belongs to the co-integrating space, a Likelihood Ratio (LR) test is performed whenever the null hypothesis of no co-integration is rejected. This test statistic follows a χ^2 distribution with $3 - r$ degrees of freedom.

4.2 Empirical analysis

We first evaluated whether prices and nominal exchange rates are integrated of order one. Our results support this conclusion for nominal exchange rates, but not for price indices, in particular the CPI, which seems to possess at least two unit roots for many countries, including the United States. This is particularly relevant in our analysis, since CPI in the United States is present in all bilateral RERs analyzed. The PPI seems to be less sensitive to this fact, with price indices appearing to be largely characterized by a single unit root, and, in some cases, even stationary. Our results below should be interpreted with some caution, having all these qualifications in mind. In particular, we expect them to be much more reliable when PPI is used.

Table 2 reports the co-integration results. Irrespective of the price index, evidence suggests that prices and the nominal exchange rate are co-integrated for most countries, but the hypothesis that the PPP vector is an element of the co-integrating space is clearly rejected. At a significance level of 5% the trace test rejects the null of no co-integration for 14 countries, for both price indices. The maximum eigenvalue test yields slightly weaker results, with 11 rejections for CPI, and 13 rejections for PPI. This agrees with other findings in the literature (for example Kugler and Lenz 1993; Culver and Papell 1999), except that our data seem to impose stronger rejections on the proportionality and symmetry conditions of the PPP hypothesis.[4] In fact, the restriction imposed by PPP on the co-integrating vector is always rejected for all cases where

[4] Baum et al. (2001) also find significant evidence for co-integration, but their results clearly reject the strict version of PPP, much in line with our conclusions.

the co-integration tests supports $r \geq 1$. Hence, according to our dataset, only a weak variant of PPP, which does not impose any restriction on the co-integrating space, can be supported.[5] The choice of the price index, again, does not induce substantial differences in the results.

4.3 An account of the results

The only evidence co-integration has brought to us was that prices and the exchange rate might be linked together in the long-run, but the proportionality and symmetry conditions that characterize the theoretical PPP vector cannot support this linkage. Some authors argue that this may be the result of measurement errors, transport costs, trade restrictions or imperfect competition, all of which create a wedge between the theoretical and the estimated co-integrating vectors, so that only a weak variant of PPP can be sustained by the data (Taylor 1988; Cheung and Lai 1993). We think, however, that these problems cannot justify the large disparities we have found among the different co-integration vectors, as, for example, (1,−27.05,35.25) for Austria, or (1,−1.90,−2.84) for Canada. In fact, the elements of these vectors may not only present the wrong signs as compared with the theoretical parameters, but may also display contradicting signs among themselves, making any attempt to interpret such estimates fruitless (Froot and Rogoff 1995). For this reason, we think that supporting a weak version of PPP which only requires prices and the exchange rate to be co-integrated makes no economic sense.

5 Panel unit root/stationarity tests

We start by presenting three standard panel unit root tests which do not take into account cross-sectional dependence between countries—the Levin et al. (2002) (henceforth LLC), the Im et al. (2003) (hereinafter IPS) and the Fisher-ADF test (Maddala and Wu 1999). The Hadri test (Hadri 2000), a panel stationarity test, is introduced in the sequel. Afterwards, we present the methodology developed in Bai and Ng (2004), known as PANIC, to tackle contemporaneous correlation in the data. We then work on our empirical results and critically access their usefulness.

5.1 Theoretical background

The main idea behind all panel unit root tests is to pool cross-section data in order to generate more powerful tests. This is a simple way to overcome one of the main critiques made to univariate unit root tests; however, as we emphasize later, it may generate additional problems not present in univariate tests, namely contemporaneous correlation between observations.

[5] Changing the lag selection criteria introduces slight changes in the results for some countries, but the overall analysis remains unchanged.

Table 2 Results of Johansen's co-integration tests

Country	r	CPI				PPI			
		p	$\lambda_{\mathbf{T}}$	$\lambda_{\mathbf{M}}$	χ^2_2	p	$\lambda_{\mathbf{T}}$	$\lambda_{\mathbf{M}}$	χ^2_2
Australia	0	3	29.70^a	20.04^a	11.65^c	1	43.13^c	34.41^c	24.06^c
	1		9.66	6.63			8.72	7.84	
Austria	0	4	40.79^c	24.63^b	17.31^c	4	20.23	9.45	0.94
	1		16.16^b	10.84			10.77	8.04	
Belgium	0	4	73.47^c	52.12^c	44.72^c	1	37.58^c	25.17^b	8.58^b
	1		21.36^c	16.60^b			12.40	12.03	
Canada	0	3	30.01^b	26.05^c	20.60^c	1	41.46^c	32.61^c	23.79^c
	1		3.96	3.79			8.85	8.84	
Denmark	0	3	44.19^c	33.02^c	17.77^c	4	32.65^b	21.23^b	11.17^c
	1		11.17	10.91			11.42	10.49	
Finland	0	4	33.36^b	19.54^a	8.95^b	1	41.09^c	32.64^c	24.65^c
	1		13.82^c	10.09			8.45	8.30	
France	0	3	28.82^a	17.52	5.83^a	n.a.	n.a.	n.a.	n.a.
	1		11.30	10.99			n.a.	n.a.	
Greece	0	4	28.51^a	16.01	8.20^b	n.a.	n.a.	n.a.	n.a.
	1		12.51	9.64			n.a.	n.a.	
Ireland	0	3	33.11^b	18.90^a	6.12^b	6	79.41^c	71.42^c	62.64^c
	1		14.21^a	11.46			7.99	7.39	
Italy	0	3	36.49^c	28.18^c	19.77^c	0	100.21^c	86.77^c	58.11^c
	1		8.31	8.27			13.44^a	11.21	
Japan	0	4	34.22^b	21.68^b	13.51^c	5	20.42	14.96	8.46^b
	1		12.54	12.46^a			5.46	4.81	
Luxembourg	0	4	56.45^c	38.31^c	31.88^c	1	25.89	16.58	8.20^b
	1		18.14^b	13.54^a			9.30	7.54	
Netherlands	0	3	28.64^a	15.47	1.18	1	36.91^c	20.42^a	9.94^c
	1		13.18	7.27			16.49^b	11.88	
New Zealand	0	3	24.25	15.32	6.61^b	1	49.44^c	31.27^c	19.72^c
	1		8.93	8.71			18.17^b	16.31^b	
Norway	0	3	36.86^c	25.22^b	14.60^c	0	58.00^c	45.67^c	23.54^c
	1		11.64	10.04			12.33	12.27	
Portugal	0	3	44.96^c	37.17^c	25.53^c	n.a.	n.a.	n.a.	n.a.
	1		7.78	6.29			n.a.	n.a.	
Spain	0	4	31.04^b	22.89^b	15.63^c	6	34.95^b	26.98^c	20.43^c
	1		8.15	7.64			7.96	7.61	
Sweden	0	3	23.73	13.32	4.11	1	40.47^c	23.34^b	8.23^b
	1		10.41	10.18			17.13^b	14.47^b	

Table 2 continued

Country	r	CPI				PPI			
		p	$\lambda_{\mathbf{T}}$	$\lambda_{\mathbf{M}}$	χ_2^2	p	$\lambda_{\mathbf{T}}$	$\lambda_{\mathbf{M}}$	χ_2^2
Switzerland	0	4	38.71^c	19.69^a	8.91^b	6	34.93^b	24.71^b	16.87^c
	1		19.03^b	12.99^a			10.22	7.14	
United Kingdom	0	4	58.13^c	27.58^c	18.35^c	4	104.72^c	87.95^c	51.16^c
	1		30.55^c	26.04^c			16.77^b	16.39^b	

Notes: a, b, and c represent rejections at 10, 5, and 1% significance levels, respectively. Co-integration tests consider an intercept and a linear trend in series, but no trend in the co-integrating equation. p denotes the number of lags included in the VEC. Lag selection was made according to the Hannan–Quinn (HQ) criterion (an upper bound of 8 lags was imposed). $\lambda_{\mathbf{T}}$ and $\lambda_{\mathbf{M}}$ denote the trace test and the maximum eigenvalue test. r is the number of co-integrating relationships considered under the null hypothesis. χ_2^2 is the test statistic on the co-integrating vector (1,1,−1). The critical values for the co-integration tests are, at 10, 5, and 1% significance levels (MacKinnon et al. 1999): 27.07, 29.80 and 35.46 ($r = 0$, $\lambda_{\mathbf{T}}$), 13.43, 15.49, and 19.94 ($r = 1$, $\lambda_{\mathbf{T}}$), 18.89, 21.13, and 25.86 ($r = 0$, $\lambda_{\mathbf{M}}$), and 12.30, 14.26, and 18.52 ($r = 1$, $\lambda_{\mathbf{M}}$). The critical values of χ_2^2 are: 4.61, 5.99, and 9.21 at 10, 5, and 1% significance levels, respectively

In order to distinguish the cross-section units, an additional index i is introduced in the logarithm of the RER. Hence, here and below the stochastic process followed by each of the N cross-section units is denoted by e_{it}.[6]

5.1.1 The LLC test

Our first panel unit root test is due to Levin et al. (2002). Here the reliance is placed on the null of a common unit root process throughout the cross-section units. The test is based on the following ADF specification

$$\Delta e_{it} = \eta_{0i} + \alpha_i e_{it-1} + \sum_{j=1}^{p_i} \eta_{ij} \Delta e_{it-j} + u_{it} \tag{12}$$

In this test equation, individual intercepts are allowed,[7] and so is heterogeneity across the cross-section dimension of the panel. Additionally, the order of serial correlation can vary freely across countries. The error term, u_{it}, is assumed i.i.d. for each i over T, but heteroskedasticity can be present across individuals.

The test is based on a modified t-statistic from a pooled regression which involves standardized variables, as a means to build the decision rule, and evaluates the null hypothesis that each RER pair follows a random walk (i.e., $\mathrm{H}_0 : \alpha_i = 0$, for every i) against the alternative hypothesis that all RER pairs are stationary and mean-revert at the same rate (i.e., $\mathrm{H}_1 : \alpha_i = \alpha < 0$, for every i). Under the null hypothesis, the

[6] Although the theoretical background presented here focuses on a balanced panel, the tests can generally be adapted and extended to the case of an unbalanced panel. See the references below for further details.

[7] A general specification should also allow for individual specific time trends. However, the inclusion of a time trend in our model violates PPP and is therefore omitted from the test equation and from the analysis.

test statistic weakly converges to a standardized normal distribution. One of the main hindrances of this inference process lies in the hypothesis to be tested. The assumption that the same first-order partial autocorrelation is shared by all the N elements of the panel under the alternative hypotheses is quite vexatious—in fact, it may even be too strong to be employed in any empirical application (Maddala and Wu 1999).

5.1.2 The IPS test

Im et al. (2003) develop a more flexible panel unit root testing procedure which avoids the unrealistic assumption of the LLC test. In particular these authors consider the same ADF specification as in Eq. 12, but allow the autoregressive coefficient to vary across the N cross-section units of the panel. The null hypothesis then becomes $\mathrm{H}_0 : \alpha_i = 0,$ for every i, which is tested against the alternative

$$\mathrm{H}_1 : \begin{cases} \alpha_i < 0, & i = 1, \ldots, N_1 \\ \alpha_i = 0, & i = N_1 + 1, \ldots, N \end{cases}$$

Rejection of the null hypothesis in our context simply means that only N_1 RER pairs are stationary, while the remaining $(N - N_1)$ pairs still have a unit root. Hence, the rejection of the null hypothesis is completely uninformative about the identity of the countries in which PPP holds, or even about the number of countries in which we expect to observe a stable RER. The misinterpretation of the null hypothesis in the IPS test is not uncommon in the literature (Sarno and Taylor 2001), and often leads to wrong conclusions.

Instead of obtaining a test statistic for panel data, Im et al. (2003) construct a properly standardized t-bar statistic by taking the average of the ADF t-statistics computed for each element of the panel. Then, through the Lindberg–Levy central limit theorem, they show that this standardized t-bar statistic converges to a standardized normal distribution under the null hypothesis. From the practical stance, it can be advantageous to consider the $W_{t\mathrm{bar}}$ statistic, which uses more reasonable choices for the mean and the variances used in the standardization process, and hence it is able to engineer a more suitable standardization of the t-bar statistic.

5.1.3 The Fisher-ADF test

The Fisher-ADF test was developed by Maddala and Wu (1999), and Choi (2001), although its roots lie in the study of Fisher (1932). This test is built under more general assumptions than the IPS test, and consequently it can perform more reliably in empirical analysis.

Consider again the process in (12), with the difference that each cross-section element of the panel can have distinct sample sizes. The null and the alternative hypotheses are the same as in the IPS test. The main idea of the Fisher-ADF test lies in combining the P values P_i of individual unit root tests for each cross-section unit, as a means to attain extra power. The mechanics of the test consist in the simple observation that, if the P values for the i-th cross-section unit are uniformly distributed in (0, 1), then $-2 \log P_i$ is chi-square distributed with two degrees of freedom. Since chi-square variables are additive, we can sum up the P values for the several cross-section units to obtain

$$P_{\lambda} = -2\sum_{i=1}^{N} \log P_i, \tag{13}$$

which is chi-square distributed with $2N$ degrees of freedom. The Fisher-ADF test is non-parametric and exact, unlike the IPS test which is parametric and only asymptotically accurate.

5.1.4 The Hadri test

The Hadri test arises as a natural extension of the KPSS stationarity test for panel data. As Hadri (2000) points out, the classical formulation of unit root tests, which evaluate the null of a unit root against the alternative of stationarity, may lead to false non-rejections unless evidence against the null is sufficiently strong. The goal here is to develop a simple residual based lagrange multiplier test, in the style of the KPSS test, in order to test the null that e_{it} is stationary for all i, against the alternative of a random walk. The simplest implementation of the test involves the estimation of a panel version of regression (10), i.e.,

$$e_{it} = \mu_i + u_{it} \tag{14}$$

where u_{it} denotes an error term. Under the alternative hypothesis, all series in the panel follow a random walk. Hence, like in the LLC test, care is needed when interpreting a possible rejection of the null, since the underlying hypothesis may be too strong to provide any valid conclusions as regards to PPP.

5.1.5 A PANIC attack on unit roots

As we noted before, ignoring cross-sectional dependence may lead to serious size distortions. In particular, the size of the panel unit root tests presented above may be substantially higher than the nominal level, inducing over-rejections of the null hypothesis (O'Connell 1998). In order to tackle this issue, we introduce the factor structure approach developed by Bai and Ng (2004), known as PANIC—Panel Analysis of Nonstationarity in the Idiosyncratic and Common components. As its name suggests, this approach is based on the decomposition of each panel series into a pervasive component and an individual specific component.[8] To our knowledge, there are very few applications of this method to PPP. One exception is given by the empirical study of Wagner (2008). This test procedure has two main advantages. First, it allows the identification of the source of non-stationarity, if any. Second, it permits the design

[8] Moon and Perron (2004) and Phillips and Sul (2003) also adopt a factor structure approach to tackle the consequences of contemporaneous correlation. Alternative methods have been developed in the literature, however. For instance, Chang (2002) adopts a covariance restrictions approach, while Pesaran (2007)

of a valid pooled test on the idiosyncratic components, since these are, in principle, independent across i.

Although the details are too complicated to be explained in detail here, the idea is simple and intuitive. Suppose that each series can be decomposed into two components—one which is mainly unit specific and one which is strongly correlated with other series. While the first may represent country specific events, such as poor economic performances which are mainly idiosyncratic, the second may capture common components that affect the worldwide economy, such as an oil crisis or depreciations of the dollar. The model can thus be written as

$$e_{it} = c_i + \lambda_i' F_t + \epsilon_{it} \tag{15}$$

where c_i is an individual specific constant, F_t is a vector of common components, and ϵ_{it} is an idiosyncratic error term. These last two components are described by

$$(I - L)F_t = C(L)u_t \tag{16}$$

$$(1 - \rho_i L)\epsilon_{it} = D_i(L)\varepsilon_{it} \tag{17}$$

where $C(L) = \sum_{j=0}^{\infty} C_j L^j$ and $D_i(L) = \sum_{j=0}^{\infty} D_{ij} L^j$. By introducing a vector of common factors that affects all series in the panel through a matrix of factor loadings λ_i, this method models directly cross-sectional dependence between individuals. The number of common factors is usually selected through an information criteria, along the lines of Bai and Ng (2002), and the number of common trends is determined by PANIC, as explained below. Obviously, F_t and ϵ_{it} are not individually observed, but the heart of PANIC lies in developing unit root tests for these elements as if they were known. Additionally, if the vector F_t effectively captures all common factors, ϵ_{it} satisfies the cross-sectional independence assumption required to construct a valid panel unit root test.

More specifically, Bai and Ng (2004) start by applying principal components to the model in (15), written in first-differences. This allows them to obtain estimates for F_t and ϵ_{it}, even if these components are not individually observed. Thereafter, they run the ADF regression on the individual de-factored series

$$\Delta\widehat{\epsilon}_{it} = d_{io}\widehat{\epsilon}_{it-1} + \sum_{j=1}^{p} d_{ij}\Delta\widehat{\epsilon}_{it-j} + \text{error} \tag{18}$$

Footnote 8 continued
proposes an augmentation of the standard ADF regressions with the cross-section averages of lagged levels and first differences of individual series. However, PANIC seems to be more general and is now the most widely accepted method.

to test for an individual specific unit root. If the common factor is unique, the existence of a common trend can be evaluated by running the ADF regression

$$\Delta\widehat{F}_t = c + \delta_o \widehat{F}_{t-1} + \sum_{j=1}^{p} \delta_j \Delta\widehat{F}_{t-j} + \text{error} \tag{19}$$

The null and alternative hypothesis are the usual ones for the ADF test. Since the limiting distributions of the t-statistics for testing a unit root in these equations are the same as the DF test (Dickey and Fuller 1979), the former for the case of no constant and the latter for the case with a constant, one can use the critical values reported in MacKinnon (1996) to assess the non-stationarity of each of the components under analysis.[9] If more than one common factor exists in data, a sequential procedure is applied, which tests the null hypothesis that m common factors are non-stationary, where m is initially equal to the number of common factors. If rejected, the test is performed again, but the number of non-stationary common factors under the null hypothesis is corrected by -1, i.e., we set $m := m - 1$. Otherwise, the estimated number of stochastic trends underlying e_{it} is m. The test statistics of this sequential procedure are based on modified versions of the Q_{f} and Q_{c} statistics (denoted by MQ_{f} and MQ_{c}, respectively) originally proposed by Stock and Watson (1988).

A series is non-stationary if at least one of these two components is non-stationary. If F_t is non-stationary but ϵ_{it} is stationary, then e_{it} will be non-stationary due to a pervasive source. If the opposite is verified, i.e., ϵ_{it} is non-stationary but F_t is stationary, then the non-stationarity of e_{it} is due to a series-specific factor that cannot be endorsed in common grounds.

This framework allowed Bai and Ng (2004) to derive a valid pooled test based on the idiosyncratic error components, since these do not depend on the common factor and are therefore independent across i.[10] Letting P_i^{ϵ} denote the P value of individual ADF tests associated to Eq. 18, the following Fisher type test can therefore be used

$$P^{\epsilon} = \frac{-2\sum_{i=1}^{N} log P_i^{\epsilon} - 2N}{\sqrt{4N}} \tag{20}$$

which follows a standardized normal distribution. The motivation for this test lies along the same lines as the Fisher-ADF test presented above, except that (20) is only valid asymptotically and hence a standardization is required.

[9] The case would be different if a linear trend was included in the test equations. The critical values presented in MacKinnon (1996) would no longer be valid, and the relevant critical values would have to be obtained by simulation.

[10] In practice, this is not so straightforward, since F_t may not capture all the pervasive components underlying the panel. The term ϵ_{it} must still be assumed independent across i.

Table 3 Panel unit root tests under the assumption of cross-sectional independence

Price Index	LLC		IPS		Fisher ADF		Hadri	
	τ_ρ	$p-v$	$W_{t\text{bar}}$	$p-v$	P_λ	$p-v$	Z_{stat}	$p-v$
CPI	−1.067	0.14	−5.198	0.00	90.294	0.00	4.319	0.00
PPI	−0.001	0.50	−4.134	0.00	70.196	0.00	3.688	0.00

Notes: $p-v$ stands for p value. The number of lags was automatically selected according to the Akaike Information Criteria (an upper bound of 8 lags was imposed) in the tests which rely on lags to correct for serial correlation. Results do not change substantially under alternative lag selection criteria. For the LLC and Hadri tests, the Bartlett kernel method was used, and bandwidth selection was made according to the Newey–West criteria. If we took into consideration the Hadri heteroskedasticity consistent Z_{stat}, the conclusion of the Hadri test would remain the same. For PPI, the unbalanced panel was used. Conclusions do not change if a balanced panel is used instead

5.2 Empirical results

The results for the LLC, IPS, Fisher-ADF, and Hadri tests are presented in Table 3. The conclusions are clear: for the tests which take non-stationarity as the null hypothesis, only the LLC test fails to reject the random walk behavior of RERs, irrespective of the price index considered. This seems to contradict the results in Oh (1996), Frankel and Rose (1996), and Wu (1995), who find evidence for stationarity using panel unit root tests with similar structures to this one. However, our evidence agrees with the findings of Papell (1997) for quarterly data when the U.S. dollar is used as numeraire. This conclusion can be motivated by the strong assumptions underlying this test, described above.

The IPS and Fisher-ADF tests clearly reject the null hypothesis of a unit root for our panel. This does not mean, however, that evidence for PPP has been found—as Sarno and Taylor (1998) point out, rejecting the null hypothesis in this context may occur even if only one series is stationary. Hence, a rejection of the null hypothesis is clearly uninformative, since the only conclusion is that PPP holds for some, but not necessarily all, countries.

Finally, the Hadri test suggests a random walk behavior for all RERs in our panel. But, once again, since a common autoregressive parameter is assumed under the null hypothesis, one should not rely too much on the inferences provided by this test. Although the null and the alternative hypothesis are exactly interchanged, the Hadri test yields the same conclusion as the LLC test, and both share a similar problem.

The results above simply ignored the existence of cross-sectional dependence. However, RERs are cross-sectionally dependent by construction, since all of them contain two common components: the U.S. price index, and the value of the dollar. We now deal with the effects of this. Table 4 provides the results of PANIC.

First of all, observe that the variability of each series that is explained by the common factors is quite heterogeneous, but overall it appears to be more relevant than the variability explained by the idiosyncratic error term. This is a first insight on the importance of factor analysis in developing and understanding the co-movements of contemporaneously related series. In order to determine the number of factors, the

Table 4 The PANIC approach

Country	CPI		PPI_{17}		PPI_{12}	
	ADF^{ϵ}	R_1	ADF^{ϵ}	R_1	ADF^{ϵ}	R_1
Australia	−0.88	0.23	−0.98	0.18	−0.76	0.19
Austria	−0.74	0.06	-2.26^{b}	0.06	-2.56^{b}	0.07
Belgium	−0.17	0.06	−0.65	0.07	n.a.	n.a.
Canada	−0.85	0.80	−1.52	0.68	-1.93^{a}	0.74
Denmark	-3.42^{c}	0.06	1.05	0.05	−1.35	0.07
Finland	−1.29	0.25	−0.16	0.18	−1.60	0.17
France	0.55	0.07	n.a.	n.a.	n.a.	n.a.
Greece	−0.71	0.30	n.a.	n.a.	n.a.	n.a.
Ireland	−0.78	0.13	−0.94	0.07	n.a.	n.a.
Italy	−1.57	0.26	−1.45	0.16	n.a.	n.a.
Japan	−0.77	0.59	−0.68	0.05	-2.77^{c}	0.05
Luxembourg	0.18	0.06	-1.75^{a}	0.07	n.a.	n.a.
Netherlands	-1.67^{a}	0.16	0.02	0.04	−1.13	0.07
New Zealand	-2.94^{c}	0.20	-2.11^{b}	0.31	−0.84	0.18
Norway	-1.99^{b}	0.15	−0.74	0.19	n.a.	n.a.
Portugal	−1.17	0.22	n.a.	n.a.	n.a.	n.a.
Spain	−0.71	0.27	-2.07^{b}	0.12	−1.23	0.24
Sweden	0.12	0.31	−1.38	0.20	-3.13^{c}	0.19
Switzerland	−1.02	0.17	−1.37	0.12	-3.98^{c}	0.13
United Kingdom	−1.12	0.41	-1.70^{a}	0.35	-2.01^{b}	0.41
Number of common factors (BIC_3)	2		3		3	
$MQ_f(3)$			−6.93		−9.50	
$MQ_f(2)$	−8.67					
Number of common trends (MQ_f)	2		3		3	
$MQ_c(3)$			−7.97		−11.09	
$MQ_c(2)$	−10.26					
Number of common trends (MQ_c)	2		3		3	
Pooled test (P^{ϵ})	2.45^{c}		2.93^{c}		6.30^{c}	

Notes: ADF^{ϵ} is the ADF test on the idiosyncratic component ϵ_{it}.R_1 is a measure of the relative importance of the idiosyncratic factor in the series, and is computed as the ratio between the variance of the first difference of the idiosyncratic error term and the variance of the first difference of the series, i.e.: $R_1 = \left(\sigma^2_{\Delta\epsilon}\right) / \left(\sigma^2_{\Delta e}\right)$. The critical values for the ADF t-stat. are approximately −1.61 at 10%, −1.94 at 5%, and −2.59 at 1% significance levels, for the case of no constant. The critical values for $MQ_{c,f}(2)$ ($MQ_{c,f}(3)$) are approximately −19.9 (−28.4) at 10%, −23.5 (−32.3) at 5%, and −31.6 (−41.1) at 1% significance levels. For PPI a balanced panel was used. For PPI_{17}, we used 104 time series observations for each of the 17 countries we have information, while PPI_{12} comprises only the 12 countries for which we have all observations, from 1973:1 until 2007:4

information criterion BIC_3 presented in Bai and Ng (2002) was used.[11] This criterion

[11] Not surprisingly, among all information criteria presented in Bai and Ng (2002), BIC_3 was the most parsimonious one, selecting the lowest number of common factors.

revealed the presence of two common factors for CPI and three for PPI. As regards stationarity analysis, both the MQ_{f} and MQ_{c} statistics indicate that all common factors are non-stationary. Furthermore, unit root tests on the idiosyncratic components show that, for a large majority of countries in our dataset, specific shocks are endowed with infinite memory.

These results provide a rich insight on the validity of PPP, since they allow us to disentangle the sources of non-stationarity. According to Table 4, PPP is not supported due to non-stationary pervasive sources. In other words, it appears that worldwide shocks contribute to permanent deviations of RERs from their long-run equilibrium values. Furthermore, for most countries in our dataset, idiosyncratic shocks contribute to strengthen the random walk behavior of RERs, adding to the effects originated by common stochastic trends. Most notably, these conclusions are, in general terms, valid both for CPI and PPI, supporting the results obtained in previous sections. The panel is only invoked to separate the common factor from the idiosyncratic component; there is no panel unit root test invoked in this argument.

Finally, the pooled P^{ϵ} test corroborates the thesis that at least some—but possibly not all—of the idiosyncratic components of RERs are stationary, for both price indices, just along the same lines of the IPS and the Fisher-ADF tests presented previously. The difference is that, as the common factor was explicitly taken into account, the error components are in principle not cross-correlated across countries, which makes any inference based on this test valid. The conclusion, however, is the same as for those tests, which suggests that the practical relevance of contemporaneous correlation is not as crucial as some literature suggests (e.g., O'Connell 1998). The true relevance of PANIC lies, in fact, in the non-stationarity tests performed on the different components of RERs.

Since the information criteria may select too many factors, we conducted the same analysis setting the number of common factors to one. In this case, the P value of the ADF test on the common component is about 5.5% for CPI and 6.5% for PPI, which supports the existence of a common stochastic trend at a 5% significance level. Thus, evidence against a non-stationary common factor appears to be stronger, but not enough to reject the null hypothesis at the usual significance level. The conclusions regarding the idiosyncratic component remain, in general terms, unchanged. Hence, according to PANIC, there is little evidence supporting PPP. In effect, this conclusion is sustained by the indication of a common stochastic trend in RERs, associated to non-stationary idiosyncratic components.

5.3 An overview of the analysis

Standard panel unit root tests are introduced in the empirical literature of PPP as a means to attain significant increases in power without having to resort to long-span data that are frequently unavailable. However, the gains are modest. First, if we assume an individual root for each country, a rejection of the null hypothesis of non-stationarity is completely uninformative, in the sense that the tests only allow us to conclude that PPP is supported for some (unidentified) RER pairs. Second, if we assume a common root, so that a rejection of the null corroborates overall PPP, then we may be imposing restrictions in the data that are too strong to be of any interest, and which will possibly lead to biased conclusions. Third, pooling the data creates a problem per se—contemporaneous correlation—which, if not dealt with, can lead to wrong conclusions. Finally, there are nowadays other alternatives to panel unit root tests that allow a large increase in power—namely the DF-GLS—and are free of the complications of these tests. One may infer that standard panel unit root tests provide no help to assert the validity of PPP.

The PANIC approach provided a new perspective on PPP, not so much due to the possibility of developing a pooled test which is robust to cross-sectional dependence, but mainly because it allows us to disentangle RERs into a common factor and an idiosyncratic component. Therefore, it becomes possible to perform a test of non-stationarity on each component separately, making possible the identification of any source of non-stationarity.

6 A univariate unit root test in the nonlinear ESTAR framework

The possibility of nonlinear adjustments is a potential pitfall when testing for the presence of a random walk in RERs (Taylor 2001). However, the idea of nonlinear adjustments is not new, dating back at least to Cassel (1922), who recognized that international transaction costs may lead to significant deviations from the LOP, consequently affecting PPP. More recently, some authors have proposed theoretical models where nonlinear adjustments arise naturally from transaction costs and frictions in international arbitrage (e.g., Dumas 1992; Sercu et al. 1995).[12] Behind these models is the idea that transaction costs may create a band of inaction within which deviations from PPP do not create profitability conditions that are the basis of convergence of RERs to their long-run equilibrium. However, once this threshold is breached, profits from international trade arise and RERs become mean-reverting.

If one believes that deviations from PPP are characterized by strong nonlinearities, as these models suggest, then standard unit root tests will have very low power to reject a potentially false null hypothesis of non-stationarity. This issue is illustrated in Taylor et al. (2001), who undertook Monte Carlo simulations to analyze the power of the traditional ADF test when the true process is characterized by nonlinearities. They show that, if the true process is globally stable, but nonlinear, the standard ADF test is biased towards the non-rejection of the null hypothesis. Testing for a unit root against the alternative of a stable, but globally nonlinear, process, in an adequate framework, is therefore necessary to assert the true dimension of nonlinearities and their consequent impact on PPP. A test of this type does not only constitute a refresh to the literature of unit root tests, but also enlightens the necessity of employing a procedure which best fits the theoretical description of the process at hand.

[12] Two alternative sources of nonlinearity have been presented in the literature. Kilian and Taylor (2003) suggest that the interaction of heterogeneous agents in the foreign exchange market at the microstructural level can lead to nonlinearities in RER adjustments, while Sarno and Taylor (2001) also suggest that a nonlinear behavior may be the consequence of official interventions in the foreign exchange market. On a different perspective, Sarno et al. (2004) discuss possible causes of nonlinear behavior in the LOP.

This idea has been successfully applied by Taylor and Peel (2000) and Taylor et al. (2001), among others, who have used Granger and Teräsvirta's (1993) Exponential Smooth Transition Autoregressive (ESTAR) model to assess the presence of nonlinearities in the RER. Their results could not be clearer: not only were they able to find evidence for nonlinear mean-reversion in RERs, but also their estimates of the degree of persistence were much lower relative to what previous studies based on panel data had suggested—a real contribution towards a deeper understanding of the PPP puzzle (Rogoff 1996; Taylor 2001; Sarno and Taylor 2002).

In this section, we ascertain whether nonlinear models are in fact able to explain the long-run behavior of RERs, by undertaking an appropriate unit root test for the nonlinear ESTAR framework. We hence contribute to a more complete understanding of PPP.

6.1 Theoretical background

The STAR model has some desirable properties that are attractive in the current context. Besides being one of the simplest nonlinear models available, it is also relatively easy to interpret. Furthermore, the literature has made available a unit root test which, under the alternative, considers precisely the type of adjustment described by this model—namely the KSS test (Kapetanios et al. 2003). In the next lines we describe both the ESTAR model and the KSS unit root test. In practice, however, the KSS test is sufficient to draw conclusions on the random walk hypothesis, and only the results for this test are reported.

6.1.1 The ESTAR model

The STAR model postulates that the RER follows a process

$$
\begin{aligned}
e_t - \mu = {} & \sum_{i=1}^{p} \phi_i [e_{t-i} - \mu] \left(1 - \Phi(\gamma, e_{t-d} - \mu)\right) \\
& + \left[\sum_{i=1}^{p} \phi_i^* [e_{t-i} - \mu] \right] \Phi(\gamma, e_{t-d} - \mu) + \epsilon_t
\end{aligned}
\tag{21}
$$

where e_t is a stationary and ergodic process, $\epsilon_t \sim iid(0, \sigma^2)$ and μ is the equilibrium RER; ϕ_i and ϕ_i^* are parameters of the model to be estimated. Following Granger and Teräsvirta (1993) and van Dijk et al. (2002), we consider the transition function

$$
\Phi(\gamma, e_{t-d} - \mu) = 1 - \exp\left[-\gamma (e_{t-d} - \mu)^2\right], \ \gamma > 0 \tag{22}
$$

which captures the idea that small shocks to the RER are highly persistent while large shocks mean-revert at a faster rate. In this case, the model is named Exponential STAR or ESTAR model. The function (22) is bounded between zero and the unity, has the properties $\lim_{|e_{t-d}| \to \infty} \Phi(\gamma, e_{t-d} - \mu) = 1$ and $\lim_{e_{t-d} = \mu} \Phi(\gamma, e_{t-d} - \mu) = 0$, and allows for a symmetric adjustment of the RER for deviations below and above the equilibrium value.[13] The parameter d represents the delay with which a deviation from the long-run equilibrium μ gives a corrective shift in the RER, and for that reason it is known as the delay parameter. Finally, γ is a strictly positive parameter which measures the speed of mean-reversion, i.e., the speed of transition from one regime to another.

Equation 21 has two interpretations. On the one hand, we can associate the ESTAR model to two regimes, each corresponding to extreme values of the transition function, $\Phi(\gamma, e_{t-d} - \mu) = 0$ and $\Phi(\gamma, e_{t-d} - \mu) = 1$. According to this perspective, if the RER follows a nonlinear mean-reverting process, it would fluctuate between a random walk regime, for small deviations from the long-run equilibrium value μ, and a mean-reverting regime, for large deviations from the equilibrium value, where the transition from one regime to the other is smooth. On the other hand, the ESTAR model can be interpreted as a continuum of regimes, each associated with a different value of the transition function, perspective according to which mean-reversion is stronger the farther way the RER is from its long-run equilibrium.

Model (21) can be decomposed in a linear and a nonlinear component

$$e_t - \mu = \sum_{i=1}^{p} \beta_i [e_{t-i} - \mu] + \left[\sum_{i=1}^{p} \beta_i^* [e_{t-i} - \mu] \right] \Phi(\gamma, e_{t-d} - \mu) + \epsilon_t \qquad (23)$$

where $\beta_i = \phi_i$ and $\beta_i^* = \phi_i^* - \phi_i$. One final reparameterization allows us to write the model in a more familiar form

$$\begin{aligned} \Delta e_t &= \alpha + \rho e_{t-1} + \sum_{i=1}^{p-1} \alpha_i \Delta e_{t-i} \\ &\quad + \left[\rho^* e_{t-1} + \sum_{i=1}^{p-1} \alpha_i^* \Delta e_{t-i} \right] \Phi(\gamma, e_{t-d} - \mu) + \epsilon_t \qquad (24) \end{aligned}$$

where $\rho = \beta_1 - 1$ and $\rho^* = \beta_1^* - 1$. If RERs are nonlinearly mean-reverting, then this specification implies that e_t may be characterized by a unit root component or even an explosive behavior for small deviations from the long-run equilibrium, but must become gradually mean-reverting the larger the distance from equilibrium. In light of this discussion, while $\rho \geq 0$ is admissible, PPP is valid if $\rho^* < 0$ and $\rho + \rho^* < 0$. This last inequality is simply the stability condition for this model.

[13] An alternative transition function, also suggested in Granger and Teräsvirta (1993), is the logistic, in which case the model is termed LSTAR. However, the LSTAR model implies an asymmetric adjustment in the RER, depending on whether it is above or below the equilibrium value. Although some argue that PPP may in fact be characterized by asymmetric nonlinear adjustments (e.g., Liew 2004), here we follow Taylor et al. (2001), that there is no obvious economic reason to consider distinct adjustment speeds depending on whether the dollar is under or overvalued.

6.1.2 A unit root test in the nonlinear ESTAR framework

The inventory of tests presented until now is unable to evaluate the existence of unit roots when the true model has nonlinearities. However, testing for a unit root directly in the above framework is rather complicated. Firstly, it involves estimating the model by nonlinear least squares. Secondly, the delay lag must be appropriately selected so that the estimated model can deliver the best fit to the data. Thirdly, any unit root test in this setup requires imposing the restrictions $\rho = 0$ and $\rho^* < 0$ in (24), so that the null hypothesis of a random walk, which can be represented by $\gamma = 0$, can be evaluated against the alternative hypothesis of a globally nonlinear but stationary process, i.e., $\gamma > 0$. This approach requires some complex Monte Carlo methods, since the distribution of γ is not known. This is further aggravated by the fact that the parameter γ may be hard to estimate (Teräsvirta 1994), since its influence on the sum of squared residuals in minimal. Although the literature has largely followed this approach, there exists a more convenient method to test for the presence of unit roots in the nonlinear ESTAR framework, to which we now turn—the KSS test.

Let $e_t^D = e_t - \mu$ denote the de-meaned data and set $p = d = 1$ in Eq. 24.[14] This yields

$$\Delta e_t^D = \rho e_{t-1} + \rho^* e_{t-1}^D \left(1 - \exp\left[-\gamma (e_{t-1}^D)^2\right]\right) + u_t \tag{25}$$

The KSS test considers that the process e_t^D follows a random walk in the middle regime, and investigates whether large deviations from the long-run equilibrium are mean-reverting. If so, then the process is locally non-stationary, but may be globally nonlinear and stationary. Hence, the KSS test enforces that $\rho = 0$ in (25), and evaluates the null hypothesis of $\gamma = 0$, to be tested against the alternative of $\gamma > 0$. Since γ is not identifiable under the null, the test can be accomplished by virtue of a first-order Taylor series expansion of (25) around $\gamma = 0$, which yields

$$\Delta e_t = \delta e_{t-1}^3 + u_t \tag{26}$$

In light of this test equation, the null hypothesis of interest can be rewritten as $\delta = 0$. In the more general case where serial correlation is present, the Taylor expansion given above holds with the due modifications. For both cases, the test statistic is asymptotically equivalent, and is

$$t_{NL} = \frac{\widehat{\delta}}{\widehat{\sigma}_\delta} \tag{27}$$

where $\widehat{\delta}$ is the least squares estimator of δ in regression (26), and $\widehat{\sigma}_\delta$ denotes its standard error. This t-statistic is asymptotically distributed as a function of a standard Brownian motion $W(r)$, with $r \in [0, 1]$. Consequently, critical values must be obtained by Monte Carlo simulations.

[14] Setting $p = 1$ is for expositional simplicity only. The test immediately extends to the case of serially correlated errors by including lagged dependent variables in the test equation below.

Up until now, we have said nothing about testing for nonlinearities. In fact, estimating an ESTAR model only makes sense if there exists evidence for nonlinearities in data, and this requires some proper testing. However, these tests only make sense if the series are stationary, which may not be the case in the current application. The problem is then clear: if one runs linearity tests prior to the KSS test, the null hypothesis of a stable and linear process, assumed in the context of linearity tests, may not be correctly formulated, potentially leading to false rejections; if one performs the KSS test without testing for nonlinearities, the alternative hypothesis that the series follows a nonlinear process of the ESTAR type may not be correctly formulated. Since, the second alternative poses fewer problems, as its only consequence is a loss in power, we do not employ any linearity test in the current framework.

Table 5 The KSS test

Country		CPI	PPI
Australia		−1.55	−2.28
Austria		−2.47	−2.40
Belgium		−2.46	−2.57
Canada		−2.29	-3.09^{b}
Denmark		−1.98	−1.59
Finland		-3.06^{b}	-2.87^{a}
France		-2.86^{a}	n.a.
Greece		−1.72	n.a.
Ireland		−2.49	-2.86^{a}
Italy		-3.22^{b}	−2.48
Japan		-2.63^{c}	−2.40
Luxembourg		−2.48	-2.65^{a}
Netherlands		-2.91^{b}	−1.97
New Zealand		-3.37^{b}	−2.14
Norway		−2.12	−1.37
Portugal		−1.58	n.a.
Spain		−2.28	-2.72^{a}
Sweden		-2.94^{b}	-2.98^{b}
Switzerland		−2.60	-2.83^{a}
United Kingdom		−2.30	−2.34
Number of rejections	*a*	7	7
	b	6	2
	c	1	0

Notes: *a*, *b*, and *c* represent rejections at 10, 5, and 1% significance levels, respectively. The KSS test considers an intercept, but no trend in the data. The number of lags was selected according to the AIC criteria (an upper bound of 8 lags was imposed). Monte Carlo simulations were performed to derive the critical values for the KSS t-statistic, with a total number of 140 observations (the dimension of our time series) and 50,000 replications. We obtained (approx.) the following critical values: −2.61 at 10%, −2.88 at 5%, and −3.45 at 1% significance levels. These values are similar to the ones obtained by Kapetanios et al. (2003)

6.2 A synopsis of the evidence

The results, presented in Table 5, seem to contradict the conclusions advanced by the pioneers of the application of nonlinear methods to PPP: for the CPI, only in 30% of the countries we were able to reject the null of a random walk; for the PPI only in two countries there seems to be evidence to support nonlinear mean-reversion. Hence, there seems to exist weak empirical evidence supporting PPP, even when allowing for nonlinear adjustments in RERs.

Let us summarize the main results so far. We started by conducting a standard univariate unit root/stationarity analysis, and concluded that only a nearly efficient test (the DF-GLS test) is able to provide some support for PPP (about 60–65% of rejections of the random walk hypothesis). Afterwards, we performed a co-integration analysis, and noted that, although for a vast majority of countries the null of no-cointegration can be rejected at a 5% significance level, the theoretical PPP vector is not supported as a member of the co-integrating space. Since the estimated co-integrating vectors are not economically meaningful, we discarded PPP under this approach. Subsequently, we resorted to panel methods to obtain further insights, but concluded that the information provided by these methods is unsatisfactory, in the sense that the rejection of the random walk hypothesis cannot be used to make insightful inferences about PPP. After that, we introduced PANIC, a panel method that allowed us to decompose each series into a common factor and an idiosyncratic component, and tested each component individually for the presence of unit roots. Using this decomposition, we concluded that RERs are characterized by common stochastic trends, which are exacerbated by individual non-stationary shocks for a majority of the elements in the panel. Finally, we argued for the existence of nonlinear adjustments in RERs, and conducted a test which is robust to this possibility. Our conclusions suggest that this approach also fails to support PPP.

Our critical assessment of the conclusions obtained herein points in two directions. The first possibility is that RERs are stationary over the very long term, so that only a highly powerful test is able to reject the random walk hypothesis, when applied over a long time horizon. This does not invalidate long-run PPP, but corroborates the observation made by many authors that the half-lives of deviations from the long-run equilibrium RER are large—sometimes about 3–5 years. Furthermore, this may explain the success of the DF-GLS over alternative tests in bringing evidence in favor of PPP. However, the different conclusions provided by the DF-GLS and the KPSS tests for some countries still need to be satisfactorily addressed.

One reason for such weak evidence in favor of PPP is the extremely high volatility of the dollar in the exchange market. The value of the dollar is mainly determined by capital flows, which are 30–40 times larger than trade flows, and may quickly change the value of the dollar, eliminating all potential gains arising from international goods arbitrage. In other words, since it takes time to ship goods, and the nominal exchange rate can vary significantly in a short period of time (as compared to prices), even large deviations from equilibrium fundamentals may not entail strong corrective pressures over the RER. In addition, if the exchange rate is determined by capital flows, a large fraction of the adjustment in the PPP equation must come from prices, which are sticky. Hence, deviations from

PPP can be long-lasting. One possibility to minimize the effects of capital flows in tests of PPP is to consider another currency as numeraire, or to consider all possible bilateral RERs. If there exists evidence that currencies of smaller countries tend to be more aligned with PPP, then the choice of the dollar as numeraire may not be appropriate.

The second possibility is the existence of specific shocks associated to the economic characteristics and performances of different countries, which implies that the long-run equilibrium RER changes over time. If this is the case, assuming a constant equilibrium may lead to the under-rejection of the random walk hypothesis. This does not completely invalidate PPP, but requires a new formulation of the theory, which may comprise economic specific factors, such as productivity shocks or risk perceptions.

Finally, there is the possibility that PPP is invalid. If so, then the direction of research lies on the theoretical side, and requires the development of models which can explain why PPP may fail, and under which conditions. In this article, we have argued that there is little empirical evidence supporting PPP, suggesting that further effort is demanded in order to reconcile theory with evidence.

7 Conclusion

One of the main goals of this article is the application of unit root/stationarity tests to study the PPP hypothesis. We applied an inventory of inference procedures, viz.: univariate unit root/stationarity tests, co-integration methods, panel unit root/stationarity tests and unit root tests adapted for nonlinear frameworks. Overall, the PANIC approach provides the richest insight on the mean-reversion of RERs, since it can identify possible sources of non-stationarity. PANIC, like all the other tests performed herein, suggests that evidence for PPP is, at best, weak, irrespective of the price index considered—CPI or PPI. A satisfactory reconciliation of the theory of PPP with empirical evidence is yet to be found.

Acknowledgments We are grateful to Anthony C. Davison, Luís Catela Nunes, Paulo Rodrigues and José Tavares for discussions, and to the editor and anonymous referees for helpful suggestions and recommendations, that led to a significant improvement of this article. We also thank Pedro Pita Barros, José Faias, Francesco Franco, Iliyan Georgiev, Ana Balcão Reis, and all participants in the IRW–FEUNL (Lisbon, March 2009) and in the 3rd Meeting of the Portuguese Economic Journal (Funchal, June 2009). Carvalho acknowledges financial support from the Center of Mathematics and Applications of Universidade Nova de Lisboa, and Júlio acknowledges financial support from Fundação para a Ciência e Tecnologia. The computations for PANIC discussed in Sect. 5 were performed through a modified MATLAB code originally made available by Christophe Hurlin. The usual disclaimer applies.

References

Bai J, Ng S (2002) Determining the number of factors in approximate factor models. Econometrica 70: 191–221

Bai J, Ng S (2004) A panic attack on unit roots and cointegration. Econometrica 72:1127–1177

Baum CF, Barkoulas JT, Caglayan M (2001) Nonlinear adjustment to purchasing power parity in the post-Bretton Woods era. J Int Money Financ 20:379–399

Cassel G (1922) Money and foreign exchange after 1914. Columbia University Press, New York

Chang Y (2002) Nonlinear IV unit root tests in panels with cross-sectional dependency. J Econom 110: 261–292
Chen B (1995) Long-run purchasing power parity: evidence from some european monetary system countries. Appl Econ 27:377–383
Cheung YW, Lai KS (1993) Long-run purchasing power parity during the recent floating. J Int Econ 34: 181–192
Cheung YW, Lai KS (1998) Parity reversion in real exchange rates during the post-Bretton Woods period. J Int Money Financ 17:597–614
Choi I (2001) Unit root tests for panel data. J Int Money Financ 20:249–272
Coakley J, Kellard N, Snaith S (2005) The PPP debate: price matters! Econ Lett 88:209–213
Corbae D, Ouliaris S (1988) Cointegration and tests of purchasing power parity. Rev Econ Stat 70:508–511
Culver SE, Papell DH (1999) Long-run purchasing power parity with short run data: evidence with a null hypothesis of stationary. J Int Money Financ 18:751–768
Dickey D, Fuller WA (1979) Distribution of the estimators for autoregressive time series with a unit root. J Am Stat Assoc 74:427–431
Dumas B (1992) Dynamic equilibrium and the real exchange rate in a spatially separated world. Rev Financ Stud 5:153–180
Edison HJ (1985) Purchasing power parity: a quantitative reassessment of the 1920s experience. J Int Money Financ 4:361–372
Edison HJ (1987) Purchasing power parity in the long-run. J Money Credit Bank 19:376–387
Elliott G, Rothenberg TJ, Stock JH (1996) Efficient tests for an autoregressive unit root. Econometrica 64:813–836
Engel C (2000) Long-run PPP may not hold after all. J Int Econ 51:243–273
Fisher RA (1932) Statistical methods for research workers, 4th Edn. Oliver & Boyd, Edinburgh
Fisher EON, Park JY (1991) Testing purchasing power parity under the null hypothesis of co-integration. Econ J 101:1476–1484
Fleissig AR, Strauss J (2000) Panel unit root tests of purchasing power parity for price indices. J Int Money Financ 19:489–506
Frankel JA (1986) International capital mobility and crowding out in the U.S. economy: imperfect integration of financial markets or goods markets? In: Hafer RW (ed) How open is the U.S. economy? Lexington Books, Lexington, MA
Frankel JA, Rose A (1996) A panel project on purchasing power parity: mean reversion within and between countries. J Int Econ 40:209–225
Froot KA, Rogoff K (1995) Perspectives on PPP and long-run real exchange rates. In: Grossman G, Rogoff K (eds) The handbook of international economics. North Holland, Amesterdam
Goldberg LG, Gosnell TF, Okunev J (1997) Purchasing power parity: modeling and testing mean reversion. J Bank Financ 21:949–966
Granger CWJ, Teräsvirta T (1993) Modelling nonlinear economic relationships. Oxford University Press, Oxford
Hadri K (2000) Testing for stationarity in heterogeneous panel data. Economet J 3:148–161
Im KS, Pesaran MH, Shin Y (2003) Testing for unit roots in heterogeneous panel. J Econom 115:53–74
Jacobson T, Nessén M (2004) Examining world-wide purchasing power parity. Empir Econ 29:463–476
Jacobson T, Lyhagen J, Larsson R, Nessén M (2008) Inflation, exchange rates and PPP in a multivariate panel cointegration model. Econ J 11:58–79
Jenkins MA, Snaith SM (2005) Tests of purchasing power parity via cointegration analysis of heterogeneous panels with consuder price indices. J Macroecon 27:345–362
Johansen S (1988) Statistical analysis of cointegration vectors. J Econ Dyn Control 12:231–254
Johansen S (1991) Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. Econometrica 59:1551–1580
Johansen S (1994) The role of the constant and linear terms in cointegration analysis of non-stationary variables. Econom Rev 13:205–230
Johansen S (1995) Likelihood-based inference in cointegrated vector autoregressive models. Oxford University Press, Oxford
Johansen S, Juselius K (1990) Maximum likelihood estimation and inference on cointegration with applications to the demand for money. Oxford B Econ Stat 52:169–209
Kapetanios G, Shin Y, Snell A (2003) Testing for a unit root in the nonlinear STAR framework. J Econom 112:359–379

Kilian L, Taylor MP (2003) Why is it so difficult to beat the random walk forecast of exchange rates? J Int Econ 60:85–107
Koedijk KG, Schotman PC, van Dijk MA (1998) The re-emergence of PPP in the 1990s. J Int Money Financ 17:51–61
Kugler P, Lenz C (1993) Multivariate cointegration analysis and the long-run validity of PPP. Rev Econ Stat 75:180–184
Kuo BS, Mikkola A (2001) How sure are we about PPP? Panel evidence with the null of stationary real exchange rates. J Money Credit Bank 33:767–789
Kwiatkowski D, Phillips PCB, Schmidt P, Shin W (1992) Testing the null hypothesis of stationarity against the alternative of a unit root. J Econom 54:159–178
Levin A, Lin C, Chu CJ (2002) Unit root tests in panel data: asymptotic and finite-sample properties. J Econom 108:1–24
Liew VK (2004) Nonlinear adjustment of ASEAN-5 real exchange rates: symmetrical or asymmetrical? Econ Bull 6:1–19
Liew VK, Chong TT, Lim K (2003) The inadequacy of linear autoregressive model for real exchange rates: empirical evidence from asian economies. Appl Econ 35:1387–1392
Lothian JR, Taylor MP (1996) Real exchange rate behavior: the recent float from the perspective of the past two centuries. J Polit Econ 104:488–510
MacKinnon JG (1996) Numerical distribution functions for unit root and cointegration tests. J Appl Econ 11:601–618
MacKinnon JG, Haug AA, Michelis L (1999) Numerical distribution functions of likelihood ratio tests for cointegration. J Appl Econ 14:563–577
Maddala G, Wu S (1999) A comparative study of unit root tests and a new simple test. Oxford B Econ Stat 61:631–652
Mark N (1990) Real and nominal exchange rates in the long-run: an empirical investigation. J Int Econ 28:115–136
Meese RA, Rogoff K (1988) Was it real? The exchange rate-interest differential relation over the modern floating-rate period. J Financ 43:933–948
Michael P, Nobay AR, Peel DA (1997) Transactions costs and nonlinear adjustment in real exchange rates: an empirical investigation. J Polit Econ 105:862–879
Moon HR, Perron B (2004) Testing for a unit root in a panel with dynamic factors. J Econom 122:81–126
Nabeya S, Tanaka K (1988) Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative. Ann Stat 16:218–235
O'Connell PJ (1998) The overvaluation of purchasing power parity. J Int Econ 44:1–19
Oh KY (1996) Purchasing power parity and unit root tests using panel data. J Int Money Financ 15:405–418
Papell DH (1997) Searching for stationary: purchasing power parity under the current float. J Int Econ 43:313–332
Pesaran MH (2007) A simple panel unit root test in the presence of cross-section dependence. J Appl Econ 22:265–312
Phillips PCB (1987) Time series regression with a unit root. Econometrica 55:277–301
Phillips PCB, Sul D (2003) Dynamic panel estimation and homogeneity testing under cross section dependence. Econ J 6:217–259
Rogoff K (1996) The purchasing power parity puzzle. J Econ Lit 34:647–668
Roll R (1979) Violations of purchasing power parity and their implications for efficient international commodity markets. In: Sarnat M, Szego GP (eds) International finance and trade. Ballinger, Cambridge, MA
Sarantis N (1999) Modeling non-linearities in real effective exchange rates. J Int Money Financ 18:27–45
Sarno L (2005) Viewpoint: towards a solution to the puzzles in exchange rate economics: where do we stand? Can J Econ 38:673–708
Sarno L, Taylor MP (1998) Real exchange rates under the recent float: unequivocal evidence of mean reversion. Econ Lett 60:131–137
Sarno L, Taylor MP (2001) Official intervention on the foreign exchange market: is it effective and, if so, how does it work? J Econ Lit 39:839–868
Sarno L, Taylor MP (2002) Purchasing power parity and the real exchange rate. IMF Staff Papers 49:65–105
Sarno L, Valente G (2006) Deviations from purchasing power parity under different exchange rate regimes: do they revert and, if so, how? J Bank Financ 30:3147–3169

Sarno L, Taylor MP, Chowdhury I (2004) Nonlinear dynamics in deviations from the law of one price: a broad-based empirical study. J Int Money Financ 23:1–25
Sercu P, Uppal R, Van Hulle C (1995) The exchange rate in the presence of transactions costs: implications for tests of purchasing power parity. J Financ 50:1309–1319
Smith V, Leybourne S, Kim T, Newbold P (2004) More powerful panel data unit root tests with an application to mean reversion in real exchange rates. J Appl Econ 19:147–170
Sollis R, Leybourne S, Newbold P (2002) Tests for symmetric and asymmetric nonlinear mean reversion in real exchange rates. J Money Credit Bank 34:686–700
Stock JH, Watson MW (1988) Testing for common trends. J Am Stat Assoc 83:1097–1107
Taylor MP (1988) An empirical examination of long-run purchasing power parity using cointegration techniques. Appl Econ 20:1369–1381
Taylor MP (1995) The economics of exchange rates. J Econ Lit 33:13–47
Taylor AM (2002) A century of purchasing power parity. Rev Econ Stat 84:139–150
Taylor AM (2001) Potential pitfalls for the purchasing power parity puzzle? Sampling and specification biases in the mean reversion tests of the law of one price. Econometrica 69:473–498
Taylor MP (2003) Purchasing power parity. Rev Int Econ 11:436–452
Taylor MP (2006) Real exchange rates and purchasing power parity: mean-reversion in economic thought. Appl Financ Econ 16:1–17
Taylor MP, Peel DA (2000) Nonlinear adjustment, long-run equilibrium and exchange rate fundamentals. J Int Money Financ 19:33–53
Taylor MP, Sarno L (1998) The behavior of real exchange rates during the post-Bretton Woods period. J Int Econ 46:281–312
Taylor AM, Taylor MP (2004) The purchasing power parity debate. J Econ Perspect 18:135–158
Taylor MP, Peel DA, Sarno L (2001) Nonlinear mean-reversion in real exchange rates: towards a solution to the purchasing power parity puzzles. Int Econ Rev 42:1015–1042
Teräsvirta T (1994) Specification, estimation and evaluation of smooth transition autoregressive models. J Am Stat Assoc 89:208–217
van Dijk D, Teräsvirta T, Franses PH (2002) Smooth transition autoregressive models—a survey of recent developments. Econom Rev 21:1–47
Wagner M (2008) On PPP, unit roots and panels. Empir Econ 35:229–249
Wallace FH, Shelley GL (2006) An alternative test of purchasing power parity. Econ Lett 92:177–183
Wu Y (1995) Are real exchange rates nonstationary? Evidence from a panel-data test. J Money Credit Bank 28:54–63