Kingston University London

Economics Discussion Papers 2023-1

RE-EVALUATING WHETHER ABSOLUTE OR RELATIVE PURCHASING POWER PARITY IS BEING TESTED WHEN USING PRICE INDICES

Chris Stewart

Kingston University
London, UK

28 April 2023

Abstract

A growing body of research suggests that the literature applying unit root, stationarity and cointegration tests of long run purchasing power parity (PPP) and the law of one price (LOP) with price index data test relative PPP/LOP and not absolute PPP/LOP. We argue that such tests cannot determine when long run relative PPP/LOP is rejected and therefore are not tests of relative PPP/LOP. These are not tests of strong absolute PPP/LOP either. We contend that they are tests of weaker forms of absolute PPP/LOP and that determining which form of absolute PPP/LOP is useful in terms of, for example, establishing the form of long run flexible-price monetary exchange rate model that it implies.

Keywords: absolute purchasing power parity, relative purchasing power parity, the law of one price, price indices, unit root tests, cointegration tests, flexible-price monetary exchange rate models.
JEL codes: C32; C43; F31

Acknowledgements: The paper has benefitted from comments made by Michela Vecchi on an earlier version. Chris Stewart is responsible for all errors.

Address for correspondence: Chris Stewart: Kingston University, School of Law, Social and Behavioural Sciences, Penrhyn Road, Kingston-upon-Thames, Surrey, KT1 2EE, UK; e-mail: c.stewart@kingston.ac.uk. Orchid id: http://orcid.org/0000-0001-7893-1364.

1. Introduction

Crownover et. al. (1996) contend that previous unit root, stationarity and cointegration tests of equations based on absolute purchasing power parity (PPP) theory using price indices are tests of long run relative PPP (RPPP) and not long run absolute PPP (APPP). Subsequently, Zhang and Zou (2014) and Zhang et. al. (2022), amongst others, have reiterated this point. We argue that such unit root, stationarity and cointegration tests can only determine whether long run RPPP holds and not whether it is rejected. Hence, they are incomplete tests of long run RPPP and therefore should not be regarded as tests of long run RPPP. We argue that while such unit root, stationarity and cointegration tests cannot test the strong form of APPP they can test weaker forms of long run APPP. Hence, tests typically applied in the literature do not test RPPP or strong APPP however, they do test weaker forms of APPP. Clarifying the form of PPP that is typically tested by the literature is the first contribution of this paper. This argument also applies to testing the law of one price (LOP). We also demonstrate that modified long run flexible-price monetary approach (FLMA) exchange rate models using levels variables can be justified when weaker forms of long run APPP hold and not when RPPP holds. Hence, we argue that testing for weaker forms of long run APPP (as we suggest the literature typically does) is useful because such tests indicate whether FLMA models can be justified. This is the second contribution of our paper.

The rest of the paper is organised as follows. Section 2 outlines absolute and relative PPP theory while a discussion of the empirical literature is given in section 3. Section 4 examines whether the empirical literature has been testing long run absolute or relative PPP. Section 5 demonstrates one use of determining whether weaker forms of long run APPP hold in terms of justifying modified long run FLMA exchange rate models. The implications of long run RPPP holding for exchange rate models are considered in section 6. Section 7 concludes.

2. Absolute and relative purchasing power parity theory

The strong form of absolute PPP (APPP) is represented by (see, MacDonald, 2007):

$$S_t = \frac{P_{1,t}}{P_{2,t}} \quad \textbf{(1)}$$

where S_t is the nominal exchange rate in period t measured as the units of country 1's currency per unit of country 2's currency, $P_{1,t}$ is country 1's price level and $P_{2,t}$ is country 2's price level.

Re-arranging **(1)** suggests that the strong form of APPP (SAPPP) predicts that the real exchange rate $\left(RER_t = \frac{S_t P_{2,t}}{P_{1,t}}\right)$ equals unity and the natural logarithm (ln) of RER_t, $ln(RER_t)$, is zero. Issues including the inaccurate measurement of variables, the proportion of goods that are traded (Harrod-Balassa-Samuelson effect), tariffs and transportation costs give rise to less stringent forms of PPP that are nested within (where α, β, and γ are constant parameters):

$$S_t = \alpha \frac{P_{1,t}^{\beta}}{P_{2,t}^{\gamma}} \quad \textbf{(2)}$$

When the symmetry ($\beta = \gamma$) and proportionality ($\beta = 1$) restrictions both hold **(2)** becomes $S_t = \alpha \frac{P_{1,t}}{P_{2,t}}$, which implies $RER_t = \alpha \neq 1$ and $ln(RER_t) = ln(\alpha) \neq 0$. We will refer to this as the weak form of absolute purchasing power parity (WAPPP). When only symmetry holds **(2)** becomes $S_t = \alpha \left(\frac{P_{1,t}}{P_{2,t}}\right)^{\beta}$, which implies:

$$lnS_t = \delta + \beta ln\left(\frac{P_{1,t}}{P_{2,t}}\right), \qquad \text{where, } \delta = ln(\alpha) \tag{3}$$

We will refer to this as WAPPP without proportionality. SAPPP holds if $\delta = 0$ and $\beta = 1$ in **(3)**. When neither proportionality or symmetry hold, we have equation **(2)** that we will call WAPPP without proportionality or symmetry. The log-linear form of **(2)** is:

$$lnS_t = ln(\alpha) + \beta lnP_{1,t} - \gamma lnP_{2,t} \tag{4}$$

Differentiating **(3)** with respect to $ln\left(\frac{P_{1,t}}{P_{2,t}}\right)$ gives the elasticity of S_t to $\frac{P_{1,t}}{P_{2,t}}$, thus:

$$\frac{dlnS_t}{dln\left(\frac{P_{1,t}}{P_{2,t}}\right)} = \beta \tag{5}$$

The strong form of relative PPP (RPPP) can be expressed as $\frac{S_t P_{2,t}}{P_{1,t}} = \frac{S_{t-1} P_{2,t-1}}{P_{1,t-1}}$, which after taking natural logarithms and rearranging gives an equivalent expression for the strong form of relative PPP (SRPPP) – see MacDonald (2007, p. 43):

$$\Delta lnS_t = \Delta lnP_{1,t} - \Delta lnP_{2,t} \quad \textbf{(6)}$$

where Δ is the first difference operator such that $\Delta X_t = X_t - X_{t-1}$. Following Coakley and Snaith (2006, p. 64) equation **(6)** can be rearranged as $\frac{\Delta lnS_t}{\Delta lnP_{1,t} - \Delta lnP_{2,t}} = \frac{\Delta lnS_t}{\Delta ln\left(\frac{P_{1,t}}{P_{2,t}}\right)} = 1$, which as Δ tends to zero gives the instantaneous elasticity of S_t to $\frac{P_{1,t}}{P_{2,t}}$, thus:

$$\frac{dlnS_t}{dln\left(\frac{P_{1,t}}{P_{2,t}}\right)} = 1 \quad \textbf{(7)}$$

where d denotes infinitely small changes. Hence, if SRPPP holds the elasticity of S_t to $\frac{P_{1,t}}{P_{2,t}}$ should equal unity. Coakley et. al. (2005, p. 296) argue that this implies the proportionality restriction ($\beta = 1$) in **(3)** will hold if (S)RPPP is valid (equating **(5)** and **(7)**). If weak RPPP $\left(\Delta lnS_t = ln(\alpha^R) + \Delta lnP_{1,t} - \Delta lnP_{2,t}\right)$ is valid $\beta = 1$ will also hold in **(3)**.

3. Empirical literature

The empirical evidence generally rejects APPP holding in every period (in the short run). Taylor and Taylor (2004, pp. 141–142) state that from the mid-1970s "… it became increasingly clear that continuous PPP could not hold as nominal exchange rates were patently far more volatile than relative national price levels". Dornbusch's (1976) exchange rate overshooting theory suggests that S_t can substantially deviate from the value predicted by APPP due to price stickiness in the short run while being forced to this value in the long run. This suggests testing should focus on whether PPP holds in the long run. The above equations that model different forms of short run PPP can be adapted to long run forms by adding a stochastic error term, u_t, that depicts deviations from the equilibrium predicted by the specified form of PPP. If $u_t \sim I(0)$ the specified form of PPP holds in the long run whereas if $u_t \sim I(1)$ it does not, where $I(d)$ denotes a process that is integrated of order d.

Since the late 1970s, a variety of unit root, stationarity and cointegration tests (including those that account for structural breaks, nonlinearity and use panel data) have been applied to determine whether different forms of PPP hold in the long run. On the general conclusions from these tests Taylor and Taylor (2004, pp. 154 - 155) state "that long-run PPP may hold in the sense that there is significant mean reversion of the real exchange rate, although there may be factors impinging on the equilibrium real exchange rate through time." Similarly, Taylor (2009) summarises the results of 18 diverse studies that empirically assess long run PPP using different methods, countries and real exchange rates and concludes that there is general support for long run PPP. Employing a panel unit root test that allows for both nonlinearity and unknown structural breaks using a Fourier function Bahmani-Oskooee et al (2015a) conclude that PPP holds for 5 major oil exporting countries (Algeria, Indonesia, Norway, Saudi Arabia, and Venezuela) if not Russia. Their use of the sequential panel selection method (SPSM) facilitates identification of those countries in the panel that reject the unit root null hypothesis and those that do not. This test also accounts for cross-sectional dependence using a bootstrap procedure. Applying quantile unit root tests that can account for abrupt and smooth structural shifts Bahmani-Oskooee et al (2017) present evidence that supports PPP for all 7 Eastern European nations that they consider. Bahramian and Saliminezhad (2021) apply Fourier quantile unit root tests that allow for structural breaks to test PPP for the ASEAN-5 economies that they suggest could form a viable single currency union. They find the real effective exchange rate to be stationary for Indonesia, the Philippines, Singapore, and Thailand, if not for Malaysia. De Villiers and Phiri (2022) apply the fractional frequency flexible Fourier form unit root test that accounts for unknown structural breaks and asymmetries (to accommodate any price frictions in the adjustment process) to the logarithm of the real effective exchange rate for 14 newly industrialised countries (NICs). PPP is supported for all 14 NICs considered.

While the above literature suggests general support for long run PPP there are numerous studies where the evidence is less supportive. Bahmani-Oskooee and Hegerty (2009) survey papers that apply unit root, stationarity and cointegration tests to data for less-developed and transition economies and conclude that the evidence is mixed regarding whether long run PPP holds for these countries. Using panel unit root tests and estimated half-lives (while accounting for different exchange rate regimes) Huang and Yang (2015) assess PPP for eleven eurozone countries and four European countries that have not adopted the euro. Once again, the results are mixed. For the 11 eurozone nations they find much less support for PPP in the post-euro period compared to the pre-euro period, although there is strong support for PPP in both sub-samples for the 4 non-eurozone countries. Their results suggest that the flexibility of the nominal exchange rate is vital to facilitate adjustment to PPP. The following two papers apply Bahmani-Oskooee et al's (2013) panel unit root test that accounts for unknown structural breaks using a flexible Fourier function and nonlinearity whilst accommodating cross-sectional dependence with a bootstrap procedure and identifying the countries where $ln(RER)$ has a unit root and those where $ln(RER)$ is stationary with the SPSM. Using this test Bahmani-Oskooee et al (2013) find that PPP holds at the 10% (5%) level for 11 (6) out of 15 Latin American countries while Bahmani-Oskooee et al (2014a) suggest PPP is supported for 26 (20) out of 34 OECD nations at the 10% (5%) level. Two papers apply panel and univariate stationarity tests that can accommodate both unknown sharp structural breaks and smooth shifts with a Fourier function as described in Bahmani-Oskooee et al (2014b). Using this test Bahmani-Oskooee et al (2014b) conclude that PPP holds for 10 out of 20 African countries' effective RERs and Bahmani-Oskooee et al (2015b) find PPP is rejected for 6 out of 8 transition economies' bilateral USA dollar RERs (using CPIs). Using a smooth time-varying cointegration test that accommodates structural breaks and allows the cointegrating relation to have time-varying coefficients, Wu et al (2018) find support for PPP in only 2 of 6 G6 countries. She et al (2021) apply both standard and Fourier versions of the Kwiatkowski et al (1992), KPSS hereafter, and augmented Dickey-Fuller (ADF) tests to determine whether bilateral $ln(RER_t)$s are stationary around multiple structural breaks and/or allowing nonlinearity for Pakistan. They employ both consumer price indices (CPIs) and wholesale price indices (WPIs). Their results offer mixed support for PPP in Pakistan because it holds against some of its 21 selected trading partners and not others. Boundi-Chraki and Mateo Tome (2022) apply a variety of linear and nonlinear unit root tests using time-series and (first and second generation) panel methods to 28 OECD countries using aggregate quarterly data for 1960Q1 to 2021Q4. CPI measures prices and the sample covers both fixed and floating exchange rate periods. The results are mixed with PPP being rejected for many countries using both linear and nonlinear methods.

When PPP holds in the long run the plausibility of the time it takes to remove half of the deviation from the equilibrium real exchange rate has been assessed using half-life calculations. Crucini and Shintani (2008, p. 641) suggest the literature's consensus was that half-lives were around 3 to 5 years. This was considered as evidence against long run PPP holding. As Rossi (2005, p. 432) states "… existing point half-life deviations from PPP are difficult to reconcile with conventional explanations for the failure of short-run PPP based on price stickiness. According to Rogoff (1996, p. 654), deviations from PPP can be attributed to transitory disturbances, like financial and monetary shocks, that buffet the nominal exchange rate and translate into real exchange rate variability because of nominal price stickiness. Thus, whereas conventional explanations for the failure of PPP based on nominal prices stickiness are compatible with the enormous short-term volatility of real exchange rates, they also imply that deviations should be short-lived, because they can occur only during a time frame in which nominal wages and prices are sticky (i.e., 1–2 years). The existing point estimates imply instead that deviations are much more persistent than that. Rogoff (1996, p. 647) called this empirical inconsistency the 'PPP puzzle'." Or as Pelagatti and Colombo (2015, p. 906) state, "… the estimated persistence in real exchange rates is too high even in those cases in which mean reversion apparently holds." Further, the generally accepted rationale for the high degree of real exchange rate persistence based on aggregate data is that the price indices used tend to include a large proportion of non-traded goods.[1] However, the evidence on this classical dichotomy for persistence in real exchange rates is "limited and decidedly mixed" Crucini and Shintani (2008, p. 630).

This general evidence on whether PPP holds in the long run most likely refers to weaker forms of absolute PPP rather than SAPPP, given by **(1)**. This is because unit root tests are typically

[1] Assuming the traded goods sector has higher productivity than the non-traded sector suggests that $ln(RER_t)$, measuring deviations from PPP, will be determined by the relative productivity of two countries according to the Harrod-Balassa-Samuelson effect. Bahmani-Oskooee and Nouira (2021) assess this Productivity Bias Hypothesis for 68 countries' $ln(RER_t)$ measured against the USA dollar using the autoregressive distributed lag (ARDL) and nonlinear ARDL bounds testing procedures (allowing for symmetric and asymmetric adjustment, respectively). They find that countries with higher productivity exhibit appreciated real exchange rates in the long run for between 27 and 44 of the 68 nations (depending on specification) giving mixed support for the Productivity Bias Hypothesis.

applied to $ln(RER_t)$ with an intercept included, thereby testing whether it converges to a non-zero mean. Further, unit root tests applied to $ln(RER_t)$ typically imply that it converges to a non-zero constant when using price indices (see, Crownover et. al 1996). Because the value of the constant compatible with SAPPP is unknown when price index data are used, unit root tests cannot determine if $ln(RER_t)$ converges to the constant predicted by SAPPP. Because Crownover et. al (1996) do not distinguish strong from weaker forms of absolute (or relative) PPP they argue that the literature discussed above has been testing relative PPP.

Coakley et. al. (2005) suggest a general test of long run RPPP using equation **(3)** with an error term (u_{it}) added and extended to a panel data context with $i = 1, \dots, N$ cross sectional units, thus:

$$lnS_{it} = \delta_i + \beta_i ln\left(\frac{P_{1,i,t}}{P_{2,i,t}}\right) + u_{it} \tag{8}$$

If $\beta \equiv E(\beta_i) = 1$ in **(8)** cannot be rejected RPPP is valid, otherwise it is not. They use panel data estimators and corresponding t-tests that give valid inference when u_{it} is $I(0)$ or $I(1)$. The test is regarded as general because it does not constrain u_{it} to be $I(0)$ as is typical with cointegration and unit root tests of PPP. An important implication is that if $u_{it} \sim I(1)$ in **(8)** the RPPP hypothesis is not necessarily rejected.[2] "In a log-levels equation relating the nominal exchange rate to the national price differential, general relative PPP implies a long-run unit slope coefficient but no restrictions on the error term. Measurement errors, transaction costs

[2] They suggest that unobserved $I(1)$ factors (possibly caused by the Harrod-Balassa-Samuelson effect) may be shifting the equilibrium RER. They argue that this can cause half-lives to be biased towards zero (giving excessively large half-lives) if that equilibrium is shifting.

or limits to arbitrage in foreign exchange markets can make the latter appear observationally equivalent to a unit root sequence. In addition, real exchange rates can be subject to transitory (nominal) or permanent (real) shocks." Coakley et. al. (2005, p. 314). They continue by stating that their "… paper proposes and implements the first tests of the general relative PPP hypothesis, which posits a long-run unit elasticity of the nominal exchange rate with respect to the price differential, in a robust framework that accommodates shifts in the equilibrium level of the real exchange rate. Simply put, if general relativity holds, then in the long run and other things equal, a 1% movement in relative prices will be offset by a commensurate movement in the nominal exchange rate, and vice versa." Using data for 19 industrialised countries and 26 developing nations they conclude that the generalised RPPP hypothesis cannot be rejected.[3]

Using aggregate annual data on absolute price *levels* between 1927 and 1992 from the source *Internationaler Vergleich der Preise flit die Lebenshaltung* Crownover et al (1996) appropriately test APPP for 15 country pairings involving Canada, France, Germany, Italy, the UK, and the USA. They test equation **(3)** for cointegration and find it exists for 10 country pairings. For these 10 country pairings they test whether $\delta = 0$ and $\beta = 1$ hold jointly to determine APPP holds. They find that APPP cannot be rejected for 4 countries. Whereas for 8 countries they cannot reject the single restriction that $\beta = 1$ and interpret this as RPPP holding for these nations.

[3] They apply their tests using both consumer and producer price indices and with the exchange rate transformed into an index, so all variables are consistently measured in index form.

More recently, Zhang and Zou (2014) and Zhang et. al. (2022) argue that they test absolute (rather than relative) PPP because they use price *level* data from the Penn World Tables (PWTs) rather than price indices. Zhang and Zou (2014) find the real exchange rate of 40 of the highest GDP nations (against the USA) is stationary using panel unit root tests.[4] However, they reject APPP because the long run RER is significantly different from unity. Similarly, while they find that the nominal exchange rate and the relative price cointegrate they reject APPP because the required coefficient restrictions are rejected.[5] Using time-series unit root tests Zhang et. al. (2022) find that Spain's bilateral RERs with 18 of its main trading partners are stationary however APPP only holds for 3 of these in terms of meeting the required coefficient restrictions. They present further evidence to suggest that departures from APPP can be explained by bilateral productivity differentials.

However, Feenstra et al (2015, pp. 3154 – 3155) state that the price variable pl_gdpo reported in PWT and used by Zhang and Zou (2014) and Zhang et al (2022) is an index equal to unity in the base year for the USA. This is confirmed by inspection of the data spreadsheet for PWT 10.0, which shows that all reported price data are indices with 2017=1 (for the USA) specified as the base year, including the price measure, pl_gdpo. This suggests that the evidence presented by Zhang and Zou (2014) and Zhang et al (2022) does not address the issue about using price indices for testing APPP raised by Crownover et al (1996).[6]

[4] Because none of these tests account for cross-sectional dependence or that the alternative hypothesis of some tests is that at least one country's RER is stationary PPP may not be supported for all countries.

[5] They test whether $d_0 = 0$ and $d_1 = 1$ in the following equation $S_t = d_0 + d_1\left(\frac{P_{1,t}}{P_{2,t}}\right) + u_t$. As they do not take logs of variables, they are not testing the conventional proportionality restriction with $d_1 = 1$.

[6] See Majumder and Ray (2020) for a review of constructing price indices to provide international and intranational price comparisons based on PPP.

Imbs et al. (2005) also criticise the use of price indices when testing PPP. They consider real exchanges rates constructed with consumer price indices where the ratio of two countries' CPIs are stationary first-order autoregressive processes with different coefficients for different goods. They show that while RER remains stationary the ordinary least squares (OLS) estimates of RER persistence exhibits a positive bias. Hence, heterogeneous adjustment speeds across goods are argued to cause aggregation bias (arising by over-weighting slowly adjusting goods in price indices) that explains the high persistence (as measured by half-lives) reported in the PPP literature. A strand of subsequent research uses disaggregated price data.

Crucini and Shintani (2008) use annual data between 1990 and 2005 for 123 cities from 79 countries for a maximum of 301 highly disaggregated products and services in any year that is available from the from the Economist Intelligence Unit (EIU) Worldwide Cost of Living Survey. Using a first-generation panel unit root test (pooled across all city pairings) they reject the unit root null hypothesis for virtually all goods within the OECD, non-OECD and USA country groupings, thereby providing strong support for weak LOP. They also find the median good's half-life deviation from LOP is 12 months for non-OECD cities, 18 months for the USA and 19 months in the OECD. For non-traded goods the median half-life in the OECD and USA is 24 months which is 6 months longer than for traded goods. The corresponding difference in medians for non-OECD countries is 2 months. This is consistent with the classical dichotomy's suggestion that aggregate RER_t persistence is due to a large proportion of non-traded goods included in price data. However, their finding that around 25% of traded goods exhibit higher persistence than the median non-traded good is not consistent with the classical dichotomy. Using aggregate CPI data Crucini and Shintani (2008) report OECD and non-OECD countries' half-lives to be 14 months and 16 months, respectively. The similarity of half-lives using aggregate CPIs and microeconomic data suggests, in contrast to Imbs et al. (2005), that aggregation bias does not explain the PPP puzzle (although there is some evidence of aggregation bias for the USA). That their half-lives are much lower than the 3 to 5 years typically reported in the literature (generally calculated with sample periods starting before 1990) indicates a decline in RER persistence through time.

Bergin et al (2013) use biannual data from the EIU Worldwide Cost of Living Survey between 1990 and 2007 for the major city in 20 industrialised nations. Their time-series unit root tests that account for cross-sectional dependence suggest that 63 (7) out of 98 (30) bilateral RER_ts for individual traded (non-traded) products are consistent with LOP. Unit root tests applied to aggregated RERs, constructed using simple averages across traded (non-traded) goods, support (reject) PPP. Hence, support for long run PPP/LOP is much stronger for traded than non-traded goods and their subsequent analysis is only for traded goods. The average half-life for disaggregated (aggregated) traded goods is 1.25 (2.10) years which is consistent with the slower adjustment speeds reported for aggregated, compared to disaggregated, data in the literature. These half-lives also suggest that RER persistence has reduced over time, which is consistent with Crucini and Shintani (2008). Using a two-equation system error-correction model (ECM) for traded goods they find the speed of adjustment to equilibrium LOP (PPP) is substantially faster in the relative price (exchange rate) equation compared to the exchange rate (relative price) equation using disaggregated (aggregated) data. They conclude that exchange rates predominantly adjust to secure PPP with aggregate data while relative prices mainly adjust to restore LOP for individual products. Results from a three equation ECM system that combines both aggregate and disaggregated variables shows that deviations from PPP and LOP at the aggregate and disaggregated level, respectively, are caused by different shocks. Half-lives for both aggregate and individual goods' RERs are the same, being around two years when caused by macroeconomic (exchange rate and price) shocks. However, the half-life for disaggregated RERs is much smaller, being about one year, when caused by idiosyncratic (product specific) price shocks. They therefore argue that the literature's finding of longer half-lives in aggregate data compared to microeconomic data is due to the different types of shock affecting these two data types. They further demonstrate that a substantial part of the aggregation bias identified by Imbs et al (2005), being due to heterogenous adjustment speeds of different individual goods' prices to macroeconomic shocks, may be minimal. This is because the biases due to macroeconomic shocks that are common across goods can cancel out in aggregate. Finally, their finding that prices adjust quickly and flexibly to idiosyncratic shocks at the disaggregated level while the effects of these adjustments cancel out at the aggregate level is not consistent with the standard sticky price model of exchange rate determination.

Robertson et al (2014) use Mexican and USA price index data for highly disaggregated matched categories of 173 "individual products" (such as apples, bananas, citrus fruits, roasted coffee, watches, audio equipment, dental services, haircuts and other personal care services, toys etc) from January 1982 to February 2010.[7] They apply first generation panel cointegration methods to test WAPPP and WAPPP without proportionality finding

[7] The price data was obtained from http://www.inegi.org.mx for Mexico and www.bls.gov for the USA. The exchange rate data was taken from www.banxico.org.mx.

overwhelming support for the latter while support for the former is mixed (WAPPP is generally supported for traded goods and rejected for non-traded goods).[8]

Pelagatti and Colombo (2015) suggest an arguably more fundamental issue with using price indices to test PPP. They show that if LOP holds for the individual goods contained in a price index, PPP will only hold if the RER is a time invariant function of the vectors of prices contained in each country's index.[9] In other words, the outcome of tests for PPP using aggregate data depend on the formula used to aggregate individual goods' prices into an index. They further demonstrate that constructing the RER as the ratio of two countries' CPIs expressed in the same currency (as is typical in applied empirical work) will rarely be a time invariant function even if LOP holds. That is, "… since the PPP is the generalization of the LOP, the presence of the latter should naturally be verified in the former. In fact our results show that, even if the LOP holds, if prices are aggregated using the CPI then the PPP does not follow." Pelagatti and Colombo (2015, p. 913).[10] "As we have shown, building CPI-based real exchange rates, even if the underlying prices display mean-reversion, determines a data-generating process which is neither stationary nor integrated, invalidating traditional ADF tests of the PPP hypothesis." Pelagatti and Colombo (2015, p. 912). They further suggest that their conclusions are relevant for other price indices used in testing PPP. Simulation results based on ADF unit root and KPSS stationarity tests as well as half-lives are reported that support their arguments on the perils of using prices indices to test PPP. These results are

[8] While time-series tests assess LOP panel tests allow assessment of PPP. Terminology regarding the different forms of PPP is inconsistent in the literature. What we term WAPPP (WAPPP without proportionality) Robertson et al (2014) call strong PPP (weak PPP).

[9] See Proposition 1 in Pelagatti and Colombo (2015, p. 909) for a technical expression of this condition.

[10] If LOP does not hold at the micro level this lowers the likelihood of PPP holding compared to when LOP holds for individual prices.

further confirmed by empirical analysis based on ADF and KPSS tests using five of the most integrated European countries where PPP would most likely be expected to hold. "Probably the main implication of our results is that the use of individual prices should be preferred to aggregate prices." Pelagatti and Colombo (2015, p. 914).

Pelagatti and Colombo (2015) use highly disaggregated data from Eurostat with, for example, food categories such as bread, breakfast cereals, butter, chocolate, eggs, fresh whole milk, fruit, pork, poultry, potatoes, and sugar.[11] While based on representative individual good prices that are generally closely matched across countries there remains some variation of goods within these categories. The implication is that the greater the degree of disaggregation of price data the more valid unit root, stationarity, cointegration and half-life methods used to assess PPP/LOP will be. Nevertheless, because the Eurostat disaggregated data are price indices the issue over the form of PPP being tested remains (following Crownover et al, 1996).

Another strand of the literature considers the cross-sectional distribution of deviations of disaggregated goods from LOP to assess PPP, where PPP suggests that the average of such deviations (averaging across the whole basket of goods) should be zero. Crucini et al (2005)

---

[11] Definitions of the PPP price level data available in Eurostat can be accessed via the following link (valid on 19/10/2022): Purchasing power parities (prc_ppp) (europa.eu). In section 3.4 of the definitions accessible via the above link the following is stated. 'PPPs can refer to a single product, a product group, or the economy as a whole. In moving up the hierarchy of aggregation, the PPPs refer to increasingly complex assortments of goods and services. Thus, if the PPP for GDP ("the economy as a whole") between Sweden and Italy is 13.18 krona to the euro, it can be inferred that for every euro spent on the GDP in Italy, 13.18 krona would have to be spent in Sweden to purchase the same volume of goods and services. Purchasing the "same volume of goods and services" does not mean that exactly identical baskets of goods and services will be purchased in both countries. The composition of the baskets will vary between countries and reflect differences in tastes and cultural backgrounds, but both baskets will, in principle, provide equivalent satisfaction or utility.' If the same baskets of goods are not used in each country the same weights are not employed in constructing aggregate price indices across nations, suggesting that more aggregated versions of this price data are subject to Pelagatti and Colombo's (2015) criticism. Data can be accessed via the following link (valid on 19/10/2022): Database - Eurostat (europa.eu). To access the data, click on the "Data Browser" icon given at the start of the line describing each data set.

assess PPP in this way for a very wide range of highly matched products using absolute price levels (rather than price induces) in EU countries for 1975, 1980, 1985 and 1990.[12] They find that for most EU country-pairings with similar incomes and VATs the over-priced and under-priced products cancel out such that the average deviations from LOP across goods is generally around zero, which broadly supports PPP across EU countries between 1975 and 1990. They also find that the tradability of a good and, more importantly, the proportion of non-traded inputs used to produce a good determines the size (variance) of deviations from LOP. Crucini et al (2005, p. 736) suggest that their use of absolute price levels allows them to answer questions about the size of deviations from LOP. They argue that this has not been done by most of the literature because it uses price index data and thereby considers relative versions of PPP/LOP (see section 4). However, their analysis is confined to four years between 1975 and 1990.

Cavallo et al (2014) use daily data from October 2008 to May 2013 on over 100,000 exactly matched individual traded goods' absolute prices levels (not price indices) across 85 countries reported on internet websites of four major retailers (Apple, Ikea, H&M and Zara). They verified that internet prices are the same as those in physical stores for the goods in their sample, are representative for the chosen countries and for other large global retailers. The firms included in their analysis comprise a non-negligible share of total spending on traded goods while the industries that they cover are estimated to account for over 20% of consumer

[12] Crucini et al (2005, p. 727) state that in "… many cases we are literally talking about the same automobile, portable radio, or type of cheese" and therefore data are often at the individual good level. Crucini et al (2005, p. 725) also observe that microeconomic price data in index form is generally available across a comparatively wide range of products which contrast with absolute price level data that is typically only available for a limited number of goods. Their use of absolute price level data (from various hard copies of Eurostat surveys of retail prices in EU capital cities) for a wide range of products is a novelty of their work. However, this data is provided for only four years (1975, 1980, 1985 and 1990) so time-series testing methods are not suitable.

spending. Overall, they find that $ln(RER_t)$ for individual goods generally deviates from zero for countries not in currency unions (even when nominal exchange rates are pegged) providing evidence against LOP. Whereas LOP generally holds for goods within the eurozone currency union and often holds among countries that use the US dollar as currency. They conclude that having a common currency is the important factor in determining whether PPP holds rather than an absence of nominal exchange rate volatility.[13] They also find that most deviations from LOP happen when a good is introduced as opposed to being a result of subsequent price movements, price stickiness or changes in the nominal exchange rate.[14]

Crucini and Telmer (2020) consider the determinants of $ln(RER)$'s variation using a panel of *absolute* retail price *levels* (denominated in local currency and not in price index form) of 301 individual goods and services sold in 123 cities from 78 nations annually between 1990 and 2015. This price data from the EIU Worldwide Cost of Living Survey is said to generally reflect a typical consumer's basket of goods and cover a similar breadth of goods as CPIs. While the data provides some limited within country variation the main price variability is across countries (which generally have different currencies). Their findings include the following. First, approximately half the deviations from LOP for a particular good are due to variation across cities/countries and around half is due to variation through time. The long-term deviations from LOP (across cities/countries) are slightly greater for non-traded goods than

[13] Using the absolute price levels (in pesos) of 40 individual goods (such as a 355ml Coca-Cola can) sold across several Mexican cities at weekly frequency between 2001 and 2011 (up to 526 observations), Elberg (2016) finds half-life deviations from PPP to be between 3 to 6 weeks. This adjustment speed is much faster than generally reported in the literature and is because Elberg (2016) considers price convergence of goods sold in different cities of the same country with the same currency whereas the literature predominantly analyses deviations from PPP/LOP across countries with different currencies. This is consistent Cavallo et al's (2014) conclusions.

[14] It is argued that these results do not support pricing models that suggest price rigidity solely arises because of high price adjustment costs. It is suggested that these models should account for variable flexible price markups as well as price complementarity.

traded goods whereas the short-term time-series deviations around these means are marginally larger for non-traded goods. Higher LOP deviations are observed for good pairings involving cities that cross international borders compared to cities within a single country.[15] Second, for fixed city pairings the overwhelming percentage (over 90%) of LOP deviations are due to variation in the mean across goods with little (less than 10%) variability in the mean deviations for each good through time. Further, there is very little common variation in LOP deviations across goods with the vast majority being good specific. This latter point suggests that while the variability in the RER may be primarily due to changes in the nominal exchange rate with sticky prices at the aggregate level this is questionable at the microeconomic level given the high variability in individual goods' prices. That is, the averaging involved to aggregate price data conceals the high time-series variability in prices of individual goods. They also find that the variability in LOP deviations across goods is largest for non-traded goods while the time-series variability in LOP deviations is substantially higher for traded goods. The latter suggests that shocks to the local currency price of traded goods is comparatively large.

Conclusions that can be drawn regarding the evidence from the empirical literature on PPP/LOP include the following. First, PPP does not hold in the short run. This led to the widespread adoption of unit root, stationarity and cointegration tests to account for

[15] Antonini et al (2022) use quarterly USA price data between 2006Q1 and 2015Q4 from www.coli.org for 132 metropolitan areas (cities) of 54 relatively narrowly defined goods and services (for example, milk, eggs, bread, toothpaste, shampoo, newspaper) to assess the causes of deviations from LOP. The two measures of LOP deviation used are the magnitude of the average bilateral log difference of prices between cities and the standard deviation of this bilateral log price difference. They find that LOP deviations are positively and significantly correlated with the physical distance between cities (confirming existing results in the literature). LOP deviations are also found to be significantly smaller if the cities in the bilateral price difference are both Democratic or both Republican compared to when one is Democrat and one is Republican. They conclude that political differences in the USA is an obstacle to competition.

potentially nonstationary data and determine whether PPP holds in the long run. Second, inferences drawn from unit root and cointegration tests that use typically available aggregate price indices should be treated with caution (Pelagatti and Colombo, 2015). As most of the empirical literature on PPP use aggregate price indices this suggest that the reliability of conclusions regarding whether long run PPP holds from most of the literature is questionable. In contrast, use of highly disaggregated price index data can provide valid inference. Third, the literature that uses disaggregated price data with unit root, cointegration and half-life methods finds mixed support for (weaker forms of) PPP/LOP. There is more general (though not unambiguous) support for PPP/LOP for traded goods however PPP/LOP is typically rejected for non-traded goods. Testing methods that account for structural breaks and nonlinearities tend to find more support for PPP/LOP than those that do not.[16] While panel data techniques can also lead to greater support for PPP/LOP, second generation methods that accommodate cross sectional dependence should be adopted to avoid biased results. Panel tests where the alternative hypothesis is that at least one cross sectional unit is stationary/cointegrated should be carefully interpreted (or apply the SPSM) to avoid overstating support for PPP/LOP. Further, half-lives appear to have declined through time such that they are between 1 and 2 years based on post-1990 data rather than the generally reported 3 to 5 years typically obtained using sample periods starting before 1990. When PPP/LOP holds adjustment to equilibrium using aggregate price data is predominantly through changes in the nominal exchange rate whereas most of the adjustment is due to changing prices at the microeconomic level. Fourth, there is research that considers the distribution of cross-section deviations from LOP using disaggregated data. These analyses

[16] We would recommend that any identified structural breaks or nonlinearities are explicitly justified by real world events to guard against spurious support for PPP/LOP.

are typically applied to a very large number goods denominated in local currency (absolute price levels) of highly matched products across numerous countries with relatively few time series observations (for example, the EIU Worldwide Cost of Living Survey is available annually from 1990 for a fee).[17] Their use of absolute price level data (rather than price indices) means that they assess SAPPP, however, their methodology does not allow assessment of weaker forms of PPP/LOP. They suggest general support for strong PPP/LOP for prices denominated in the same currency (either within the same country, in a currency union or for countries that use a common currency). In contrast, strong PPP/LOP is generally rejected for goods in different locations having prices denominated in different currencies. Fifth, Crownover et al (1996) suggest that the ubiquitous use of price indices means that the literature has been testing relative PPP and not absolute PPP. We now discuss this issue in more detail.

4. Has the literature using price indices been testing absolute or relative PPP?

Crownover et. al. (1996, p. 785) argue that if price indices are used to measure price levels, equation **(1)** becomes:

$$S_t/S_B = \frac{(P_{1,t}/P_{1,B})}{(P_{2,t}/P_{2,B})} \tag{9}$$

[17] The data from the 2022 edition of the EIU Worldwide Cost of Living Survey are based on more than 400 actual prices paid for over 200 surveyed goods and services in over 170 cities around the world (Worldwide Cost of Living - Economist Intelligence Unit (eiu.com) and https://wcol.eiu.com/asp/wcol_HelpPrices.asp). Prices are converted into USA dollars using current exchange rates. Aggregated price indices are constructed using identical weights to allow international comparisons. This data is not available for free although may be available in the libraries of academic institutions.

where B represents the base period of the index. Tests of PPP using price indices are therefore based on **(9)**, which relates the change of the exchange rate in period t from the exchange rate in period B to the ratio of the changes of prices in period t from period B. Taking the natural log of both sides of **(9)**, re-arranging and adding an error term (u_t), gives:

$$ln(RER_t) = ln\left(\frac{S_t P_{2,t}}{P_{1,t}}\right) = ln\left(\frac{S_B P_{2,B}}{P_{1,B}}\right) + u_t \qquad \textbf{(10)}$$

Hence, when using price indices, $ln(RER_t)$ will no longer necessarily tend to zero in the long run when (the strong form of) absolute PPP is true. Rather it tends to the unknown constant $ln\left(\frac{S_B P_{2,B}}{P_{1,B}}\right)$,[18] since the base period terms are constant. Similarly, the weak form of PPP without proportionality using price indices is $S_t/S_B = \alpha\left[\frac{(P_{1,t}/P_{1,B})}{(P_{2,t}/P_{2,B})}\right]^{\beta}$, which after taking the natural logarithm and adding an error term (u_t), implies:

$$lnS_t = \delta + \beta ln\left(\frac{P_{1,t}}{P_{2,t}}\right) + u_t \qquad \textbf{(11)}$$

where,

$$\delta = ln(\alpha) + ln\left(\frac{S_B P_{2,B}^{\beta}}{P_{1,B}^{\beta}}\right) \qquad \textbf{(12)}$$

[18] Only if $\frac{S_B P_{2,B}^{\beta}}{P_{1,B}^{\beta}} = 1$ will $ln(RER_t)$ tend to zero in the long run, which cannot generally be assumed.

When not using price indices SAPPP holds if $\alpha = 1$. Substituting $\alpha = 1$ into **(12)** gives $\delta = ln\left(\frac{S_B P_{2,B}^{\beta}}{P_{1,B}^{\beta}}\right)$, which suggests the intercept in **(11)** will not necessarily equal zero. Hence, the use of price indices means that (strong) APPP no longer necessarily predicts that the intercept in **(11)** is zero and so this theoretical prediction cannot be empirically tested because $\frac{S_B P_{2,B}^{\beta}}{P_{1,B}^{\beta}}$ is unknown. Nevertheless, theoretical predictions about β can be tested when price indices are used.[19] Since (strong) APPP requires $\alpha = 1$ and $\beta = 1$, time-series (cointegration) tests based on **(11)** and (unit root) tests based on **(10)** cannot determine whether $\alpha = 1$ when using price indices, they can only assess if $\beta = 1$. It is argued that this implies that (the strong form of) absolute PPP cannot be tested by such tests although relative PPP, which implies $\beta = 1$, can. Crownover et. al. (1996, p. 785) therefore conclude that, "… all previous time-series tests for PPP are based on price indices and are tests for relative PPP. … Therefore previous evidence that provides support for relative PPP … provides only incomplete support for absolute PPP". They argue that absolute PPP can only be tested using data that measures price levels. Importantly, and in contrast to us, Crownover et. al. (1996) do not distinguish strong absolute (or relative) PPP from weaker forms.

Contrary to the conclusions of Crownover et. al. (1996), we argue that past unit root (stationarity) and cointegration time-series tests of PPP using (valid disaggregated) price

[19] An alternative way of viewing this is to note that when a constant, X_B, is subtracted from a time-series variable, X_t, to give $X_t - X_B$, all that has happened is that X_t has been rescaled. The two variables X_t and $X_t - X_B$ have the same variation (the correlation coefficient between them is unity) although their means are different. Hence, estimating the following two regressions by OLS, $Y_t = a_1 + a_2 X_t + v_{1t}$ and $Y_t = a_1^* + a_2(X_t - X_B) + v_{2t}$, will give the same slope, b_2, however the intercepts (a_1 and a_1^*) will be different, given the intercept measures the difference in the means of the left and right-hand sides of an equation. The implications of using price indices to measure the level of prices should also be considered for other models, such as the money demand equations given below or any models that use price indices to create real variables.

indices are incomplete tests of long run relative PPP (RPPP) and, therefore, are not tests of RPPP.[20] While they are also not tests of strong APPP, we argue that they are tests of whether weaker forms of absolute PPP hold in the long run.

It is true that rejecting the unit root null hypothesis applied to $ln(RER_t)$ measured using price indices (based on **(10)**) confirms that $ln(RER_t)$ is continuously forced to a constant which provides support for long run (strong) RPPP.[21] However, if the unit root null hypothesis applied to $ln(RER_t)$ cannot be rejected this means that $ln(RER_t)$ is not converging to a constant mean. While this rejects long run weak APPP (WAPPP) it does not reject long run RPPP. Long run RPPP does not require $ln(RER_t)$ to be stationary, rather it is only rejected if $\Delta ln(RER_t)$ is nonstationary. Hence, unit root tests applied to $ln(RER_t)$, based on price indices, can only provide incomplete information on whether RPPP holds in the long run because they cannot determine when long run RPPP is rejected. As such, unit root tests applied to $ln(RER_t)$ are not tests of long run RPPP, contrary to Crownover et. al. (1996). However, they are tests of whether WAPPP holds in the long run.

Now consider cointegration tests applied to **(11)** when using price indices. If the null hypothesis of no cointegration is rejected *and* $\beta = 1$ this supports both WAPPP and (strong) RPPP holding in the long run.[22] If the no cointegration null hypothesis is not rejected this

[20] While tests of PPP typically use price indices to measure prices, they rarely utilise exchange rate indices which implies that $S_B = 1$ in **(9)**, and all subsequent equations where S_B appears. Although this does not change the point that the constant implied by (strong) APPP is unknown, it does clarify that it is due to an issue with data on price indices and not generally the exchange rate.

[21] If $ln(RER_t)$ is stationary around a constant mean (see **(10)**), its difference should be stationary around a zero mean, consistent with long run strong RPPP. However, this is not explicitly a test of $[ln(\alpha^R) =]\ \delta^R = 0$ in $\Delta ln(RER_t) = \delta^R$.

[22] For SRPPP this is because differencing **(11)** with $\beta = 1$ imposed implies $\Delta ln(RER_t)$ is forced towards a zero mean.

implies $u_t \sim I(1)$ in **(11)** and long run WAPPP without proportionality is rejected. However, RPPP is not necessarily rejected. Recall that Coakley et. al. (2005) employ essentially the same PPP specification (equation **(8)**) using *panel* data estimators to test general relative long run PPP allowing $u_t \sim I(1)$ or $u_t \sim I(0)$. Whether long run RPPP holds does not rely on $u_t \sim I(0)$ rather it requires the testing of $\beta = 1$. When *time-series* cointegration tests find $u_t \sim I(1)$ using **(11)** researchers do not typically continue to test $\beta = 1$ because the test will be subject to spurious regression and so it cannot be established whether long run RPPP holds. Hence, cointegration *time-series* tests applied to **(11)** are not tests of long run RPPP. Rather they test whether long run WAPPP without proportionality holds. If cointegration is found, testing whether $\beta = 1$ determines whether long run weak absolute PPP (with proportionality imposed) holds. However, additionally testing whether $\delta = 0$ does not establish whether long run strong absolute PPP (SAPPP) holds because the intercept need not be zero when price indices are used.

We now further consider why long run RPPP is not rejected when there is no cointegration, that is, $u_t \sim I(1)$ in **(11)**. Consistent with the literature assume $lnS_t \sim I(1)$ and $ln\left(\frac{P_{1,t}}{P_{2,t}}\right) \sim I(1)$. For example, for 45 countries Coakley et. al. (2005, pp. 306-307) find that the logs of the nominal exchange rate and relative price can be treated as $I(1)$. Hence, when u_t is found to be $I(1)$ in **(11)** there is no long run levels relationship between the $I(1)$ lnS_t and $ln\left(\frac{P_{1,t}}{P_{2,t}}\right)$ series and WAPPP without proportionality is rejected. However, because $\Delta lnS_t \sim I(0)$ and $\Delta ln\left(\frac{P_{1,t}}{P_{2,t}}\right) \sim I(0)$ these two series will not diverge in the long run, and so long run RPPP is not necessarily rejected when $u_t \sim I(1)$. Whether strong RPPP (SRPPP) holds in the long run depends on whether a long run equation of the form of **(13)** has an $I(0)$ residual and $\delta^R = 0$ and $\beta = 1$.[23]

$$\Delta lnS_t = \delta^R + \beta(\Delta lnP_{1,t} - \Delta lnP_{2,t}) \quad \textbf{(13)}$$

To illustrate the implications of our analysis we reinterpret Crownover et al's (1996) empirical PPP test results based on **(2)** for their 15 country pairings as follows. We interpret their finding of cointegration for 10 country pairings as supporting long run WAPPP without proportionality for these 10 pairs and rejecting it for the other 5. For 4 nation pairings they cannot reject the joint restriction of $\delta = 0$ and $\beta = 1$ placed on **(2)** holding. We interpret this as supporting long run SAPPP for these 4 pairs and not simply APPP as they conclude. That they cannot reject the single restriction that $\beta = 1$ for 8 countries we interpret as evidence in favour of long run WAPPP for these 8 nations and not just RPPP as they suggest. However, because their results use aggregate price data that cannot be assumed to use time invariant functions of the prices contained in each country's price variable measure, they are likely subject to the criticisms of Pelagatti and Colombo (2015). We therefore treat Crownover et al's (1996) results (and the conclusions based upon them) with caution.

5. Implications of weaker forms of absolute PPP for FLMA models

We now consider why it is of interest to determine whether weaker forms of long run absolute PPP (APPP) and absolute LOP hold and why PPP testing should not exclusively focus

[23] While the error term of an equation based on **(13)** may be stationary a static long run solution will need to be obtained if lags of the variables need to be added to remove autocorrelation.

on long run SAPPP. This is based our contention that these weaker forms of APPP are being assessed when applying valid unit root, stationarity and cointegration tests using (disaggregated) price indices. MacDonald (2007, p. 94) suggests that flexible-price monetary approach (FLMA) models of the exchange rate assume that strong APPP (SAPPP) holds in the short run while sticky-price monetary models only assume that SAPPP holds in the long run.[24] For simplicity, we consider FLMA exchange rate models. While these models assume APPP holds in the short run they can be viewed as long run relations. MacDonald (2007, p. 96, our comments in square parentheses) states the following. "Although some proponents of the flex[ible]-price monetary model view equation [**(20)**] as holding continuously it seems more appropriate to think of it as a long-run equilibrium relationship, where the nominal interest rates, via the Fisher condition, capture expected inflation." Viewing the standard FLMA model, given by **(20)** below, as a long run relationship implies that SAPPP is only required to hold in the long run and tests of long run APPP are useful in determining whether **(20)** is justified as a long run relation. Further, money demand equations and money market equilibria (given below) only need to hold in the long run for **(20)** to be viewed as a long run equation.

Assuming that domestic and foreign bonds are perfect substitutes means that the FLMA exchange rate model can be developed by focussing on the money market equilibrium (see, MacDonald, 2007, p. 95). Assume that country 1 and country 2's money demands can be

[24] Asset price versions of the monetary model view PPP as a relation between the exchange rate and asset prices (rather than goods prices). Since asset prices are more volatile than goods prices such a specification seeks to address the issue raised by Rogoff (1996) that exchange rates are substantially more volatile than goods prices. The asset price approach has two versions. The first is the monetary approach that assumes different countries' non-money assets (bonds) are perfect substitutes and the second is the portfolio balance approach that assume different nations' bonds are imperfect substitutes.

summarised by the following log-linear equation (where $\varphi_1 > 0$ and $\varphi_2 < 0$ are assumed to be the same in both countries):

$$lnM_{j,t}^{D} - lnP_{j,t} = \varphi_1 lnY_{j,t} + \varphi_2 lnR_{j,t}, \quad j = 1, 2 \tag{14}$$

where, $M_{j,t}^{D}$ denotes money demand for country j, $Y_{j,t}$ is income, and $R_{j,t}$ is the nominal interest rate on bonds. Assuming the money market is in long run equilibrium in both countries, $lnM_{j,t}^{D} = lnM_{j,t}^{S} = lnM_{j,t}$, where $M_{j,t}^{S}$ is country j's money supply ($M_{j,t}^{S}$) and substituting this into **(14)**, implies: $lnP_{j,t} = lnM_{j,t} - \varphi_1 lnY_{j,t} - \varphi_2 lnR_{j,t}$. Subtracting $lnP_{2,t}$ from $lnP_{1,t}$, yields:

$$lnP_{1,t} - lnP_{2,t} = \left(lnM_{1,t} - lnM_{2,t}\right) - \varphi_1\left(lnY_{1,t} - lnY_{2,t}\right) - \varphi_2\left(lnR_{1,t} - lnR_{2,t}\right) \tag{15}$$

Re-arranging the log-linear WAPPP without proportionality or symmetry equation, **(4)**, gives:

$$lnP_{1,t} - \frac{\gamma}{\beta} lnP_{2,t} = \frac{lnS_t - ln(\alpha)}{\beta} \tag{16}$$

Substitution of **(16)** into **(15)** implies the following modified long run FLMA exchange rate model that is consistent with long run WAPPP without proportionality and symmetry:

$$lnS_t = ln(\alpha) + (\beta - \gamma) lnP_{2,t} + \beta\left(lnM_{1,t} - lnM_{2,t}\right) - \beta\varphi_1\left(lnY_{1,t} - lnY_{2,t}\right) - \beta\varphi_2\left(lnR_{1,t} - lnR_{2,t}\right) \tag{17}$$

Placing the relevant restrictions on **(17)** gives modified versions of long run FLMA exchange rate models implied by stronger forms of APPP. Imposing symmetry ($\beta = \gamma$) on **(17)** yields the modified FLMA model corresponding to WAPPP without proportionality, thus:

$$lnS_t = ln(\alpha) + \beta\left(lnM_{1,t} - lnM_{2,t}\right) - \beta\varphi_1\left(lnY_{1,t} - lnY_{2,t}\right) - \beta\varphi_2\left(lnR_{1,t} - lnR_{2,t}\right) \quad \textbf{(18)}$$

Applying both symmetry and proportionality ($\beta = \gamma = 1$) to **(17)** gives the modified FLMA model implied by WAPPP, being:

$$lnS_t = ln(\alpha) + \left(lnM_{1,t} - lnM_{2,t}\right) - \varphi_1\left(lnY_{1,t} - lnY_{2,t}\right) - \varphi_2\left(lnR_{1,t} - lnR_{2,t}\right) \quad \textbf{(19)}$$

Imposing the restrictions implied by SAPPP ($\alpha = \beta = \gamma = 1$) on **(17)** yields the following standard FLMA exchange rate model:

$$lnS_t = \left(lnM_{1,t} - lnM_{2,t}\right) - \varphi_1\left(lnY_{1,t} - lnY_{2,t}\right) - \varphi_2\left(lnR_{1,t} - lnR_{2,t}\right) \quad \textbf{(20)}$$

Therefore, when long run SAPPP holds the standard FLMA exchange rate model given by **(20)** is justified. Weaker forms of APPP lead to modified versions of the long run standard FLMA exchange rate model. When long run WAPPP or long run WAPPP without proportionality hold the implied exchange rate models (**(18)** and **(19)**) include the same variables as the standard FLMA equation although (some of) the coefficients change. The modified long run FLMA exchange rate model adds lnP_t^* to the variables included in the standard FLMA equation when long run WAPPP without proportionality and symmetry holds (see **(17)**). Hence, tests of weaker forms of APPP (and not only SAPPP) are useful for determining whether (modified)

FLMA exchange rate models can be justified. Further, all these modified FLMA specifications are nested within the following linear model and all theoretical parameters can be identified from its OLS (or nonlinear) estimated coefficients.

$$lnS_t = b_1 + b_2 lnP_{2,B} + b_3(lnM_{1,t} - lnM_{2,t}) + b_4(lnY_{1,t} - lnY_{2,t}) + b_5(lnR_{1,t} - lnR_{2,t}) \quad \textbf{(21)}$$

where, $\alpha = exp\ (b_1)$, $\beta = b_3$, $\gamma = \beta - b_2$, $\varphi_1 = -\frac{b_4}{\beta}$, $\varphi_2 = -\frac{b_5}{\beta}$.

Hence, whether the long run FLMA model can be justified, and if it can in which form, depends (in part) on the form of long run absolute PPP that is supported by the data. Tests of whether weaker forms of absolute PPP/LOP hold in the long run therefore provide useful information. Such tests can be validly implemented provided sufficiently disaggregated price index (or absolute price level) data or appropriately aggregated price indices are used so that they avoid the criticisms raised by Pelagatti and Colombo (2015). There have been valid assessments of SAPPP using disaggregated absolute price level data in local currency. However, they are small in number, typically use a relatively short time-series dimension and often have a limited coverage of goods. Nevertheless, if base year values of the price levels ($P_{1,B}$ and $P_{2,B}$) for (disaggregated) price index data were published by data providers this would facilitate more widespread testing of long run SAPPP using price indices with larger sample sizes.

6. Implications of relative PPP for exchange rate models

Weak relative PPP (WRPPP) without proportionality and symmetry may be expressed as, $\Delta lnS_t = ln\alpha^R + \beta^R \Delta lnP_{1,t} - \gamma^R \Delta lnP_{2,t}$, where α^R, β^R, and γ^R are constant parameters. This can be rearranged as:

$$\Delta lnP_{1,t} = \frac{\Delta lnS_t - ln\alpha^R}{\beta^R} + \frac{\gamma^R}{\beta^R}\Delta lnP_{2,t} \quad \textbf{(22)}$$

Differencing **(15)** gives:

$$\Delta lnP_{1,t} - \Delta lnP_{2,t} = \Delta\left(lnM_{1,t} - lnM_{2,t}\right) - \varphi_1\Delta\left(lnY_{1,t} - lnY_{2,t}\right) - \varphi_2\Delta\left(lnR_{1,t} - lnR_{2,t}\right) \quad \textbf{(23)}$$

Substitution of **(22)** into **(23)** implies:

$$\Delta lnS_t = ln(\alpha^R) + (\beta^R - \gamma^R)\Delta lnP_{2,t} + \beta^R\Delta\left(lnM_{1,t} - lnM_{2,t}\right) \quad \textbf{(24)}$$
$$-\beta^R\varphi_1\Delta\left(lnY_{1,t} - lnY_{2,t}\right) - \beta^R\varphi_2\Delta\left(lnR_{1,t} - lnR_{2,t}\right)$$

When long run WRPPP without proportionality and symmetry holds it implies the long run exchange rate model given by **(24)**. Imposing symmetry ($\beta^R = \gamma^R$) on **(24)** gives the exchange rate model corresponding to WRPPP without proportionality, thus:

$$\Delta lnS_t = ln(\alpha^R) + \beta^R\Delta\left(lnM_{1,t} - lnM_{2,t}\right) - \beta^R\varphi_1\Delta\left(lnY_{1,t} - lnY_{2,t}\right) \quad \textbf{(25)}$$
$$-\beta^R\varphi_2\Delta\left(lnR_{1,t} - lnR_{2,t}\right)$$

Applying both symmetry and proportionality ($\beta^R = \gamma^R = 1$) to **(24)** gives the exchange rate model implied by WRPPP, being:

$$\Delta lnS_t = ln(\alpha^R) + \Delta\left(lnM_{1,t} - lnM_{2,t}\right) - \varphi_1\Delta\left(lnY_{1,t} - lnY_{2,t}\right) - \varphi_2\Delta\left(lnR_{1,t} - lnR_{2,t}\right) \quad \textbf{(26)}$$

Imposing $\alpha^R = \beta^R = \gamma^R = 1$ on **(24)** yields the exchange rate model implied by strong relative PPP (SRPPP), thus:

$$\Delta lnS_t = \Delta\left(lnM_{1,t} - lnM_{2,t}\right) - \varphi_1\Delta\left(lnY_{1,t} - lnY_{2,t}\right) - \varphi_2\Delta\left(lnR_{1,t} - lnR_{2,t}\right) \quad \textbf{(27)}$$

Equations **(24)** to **(27)** show that if any form of long run RPPP is valid the exchange rate models implied by RPPP involve relations between differenced variables. Long run RPPP does *not* justify any long run FLMA exchange rate model in *levels* variables whereas long run APPP holding does. This illustrates the importance of distinguishing tests of RPPP from tests of different forms of APPP because the implications are different. Being able to test for weaker forms of APPP is useful because it indicates whether, for example, modified FLMA exchange rate models are supported. Hence, it is important to know that typical applications of unit root (stationarity) and cointegration methods to PPP using highly disaggregated price indices (price levels) or appropriately aggregated price indices are valid tests of weaker forms of APPP (or absolute LOP) and not RPPP (or relative LOP).

For example, Crownover et al's (1996) interpretation of their evidence was that long run RPPP is supported for 8 out of the 15 country pairings that they consider and is rejected for the remaining 7. Their interpretation does not support the use of (modified) FLMA exchange rate models for any of these country pairings. Whereas our interpretation of their results is that some form of long run APPP and, therefore some form of modified FLMA exchange rate model, are supported for the 10 country pairings where cointegration is found. Further, for the 4 nation pairings where the joint restriction of $\delta = 0$ and $\beta = 1$ placed on **(2)** cannot be rejected both SAPPP and the standard FLMA exchange rate models are supported in the long run.

7. Conclusions

The following conclusions can be drawn. First, while most assessments of PPP conducted in the literature using aggregate price indices are invalid those that use sufficiently disaggregated price indices or absolute price level data expressed in currency are arguably valid (Pelagatti and Colombo, 2015). When using valid price index data, typical unit root (stationarity) and cointegration tests do not test long run strong absolute PPP/LOP (APPP/ALOP). Nor do they test long run relative PPP/LOP (RPPP/RLOP) because they cannot indicate when long run RPPP/RLOP is rejected, they can only determine when it is not rejected. Rather, these are tests of weaker forms of long run APPP/ALOP. Second, all forms of long run APPP justify (modified) versions of long run FLMA exchange rate models involving levels of the variables. Different forms of long run APPP have different implications for the parameters in the derived exchange rate equation while long run WAPPP without either proportionality or symmetry imposed suggests a FLMA model modified by adding the foreign price variable. In contrast, if long run RPPP holds it does not imply a long run FLMA exchange rate model involving levels variables is justified. Hence, it is crucially important to be able to identify if tests determine whether long run APPP or long run RPPP holds and, in the former case, in what form. We are not aware of these two points being made previously and they represent the contributions of this paper.

The significance of our paper is threefold. First, past research using unit root (stationarity) tests applied to $ln(RER_t)$ and cointegration tests between lnS_t and $ln\left(\frac{P_{1,B}}{P_{2,B}}\right)$ or between lnS_t, $lnP_{1,B}$ and $lnP_{2,B}$ that use valid price index data should now be viewed as tests of weaker forms of long run APPP rather than tests of long run RPPP. Second, we recommend that in future researchers view tests of weaker forms of long run APPP as worthwhile because they are still useful in, for example, justifying modified versions of long run FLMA exchange rate models. Third, future research should consider the implications of weaker forms of APPP for other exchange rate specifications.

References

Antonini, M., Fielding, D. and Pires, J. (2022) "E pluribus unum? Political fractionalization and deviations from the law of one price in the USA", *Oxford Economic Papers*, 74(1), 194–214.

Bahmani-Oskooee, M., Chang, T., Cheng, S. C. and Tzeng, H. (2017) "Revisiting purchasing power parity in Eastern European countries: quantile unit root tests", *Empirical Economics*, 52, 463-483.

Bahmani-Oskooee, M., Chang, T., Cheng, S. C. and Wu, T. (2015a) “Revisiting purchasing power parity in major oil-exporting countries”, *Macroeconomics and Finance in Emerging Market Economies*, 8(1-2), 108-116.

Bahmani-Oskooee, M., Chang, T. and Hung, K. (2013) “Revisiting purchasing power parity in Latin America: Sequential panel selection method”, *Applied Economics*, 45, 4584–90.

Bahmani-Oskooee, M., Chang, T. and and Liu, W. (2014a) “Revisiting purchasing power parity in 34 OECD countries: sequential panel selection method”, *Applied Economics Letters*, 21(18), 1283-1287.

Bahmani-Oskooee, M., Chang, T. and Wu, T. (2014b) “Revisiting purchasing power parity in African countries: panel stationary test with sharp and smooth breaks”, *Applied Financial Economics*, 24(22), 1429-1438.

Bahmani-Oskooee, M., Chang, T. and Wu, T. (2015b) “Purchasing Power Parity in Transition Countries: Panel Stationary Test with Smooth and Sharp Breaks”, *International Journal of Financial Studies*, 3, 153-161.

Bahmani-Oskooee, M. and Hegerty, S. H. (2009) “Purchasing Power Parity in Less-Developed and Transition Economies: A Review Paper”, *Journal of Economic Surveys*, 23(2), 617–658.

Bahmani-Oskooee, M. and Nouira, R. (2021) “The nonlinear ARDL approach and productivity bias hypothesis: Evidence from 68 countries”, *The Quarterly Review of Economics and Finance*, 80, 80–89.

Bahramian, P. and Saliminezhad, A. (2021) “Revisiting purchasing power parity in the ASEAN-5 countries: evidence from the Fourier quantile unit root test”, *Applied Economics Letters*, 28(9), 1104-1109.

Bergin, P. R., Glick, R. and Wu, J. L. (2013). “The micro-macro disconnect of purchasing power parity”, *The Review of Economics and Statistics*, 95(3), 798–812.

Boundi-Chraki, F. Mateo Tomé, J. P. (2022) “The purchasing power parity hypothesis tested once again. New empirical evidence for 28 OECD countries”, *Investigación Económica*, 81(322), 3-26.

Cavallo, A., Neiman, B. and Rigobon, R. (2014). “Currency unions, product introductions, and the real exchange rate” *The Quarterly Journal of Economics*, 529–595.

Coakley, J., Flood, R., Fuertes, A. and Taylor, M. (2005) “Purchasing power parity and the theory of general relativity: the first tests”, *Journal of International Money and Finance*, 24, 293–316.

Coakley, J. and Snaith, S. (2006) “Testing for symmetry and proportionality in a European panel”, *Applied Economics Letters*, 16, 63–71.

Crownover, C., Pippenger, J., Steigerwald, D. G. (1996) “Testing for absolute purchasing power parity”, *Journal of International Money and Finance*, 15, 783–796.

Crucini, M. and Shintani, M. (2008). “Persistence in law of one price deviations: evidence from micro-data”, *Journal of Monetary Economics*, 55, 629–644.

Crucini, M. and Telmer, C. (2020). “Microeconomic sources of real exchange rate variation”, *Review of Economic Dynamics*, 38, 22–40.

Crucini, M., Telmer, C. and Zachariadis, M. (2005). “Understanding European real exchange rates”, *American Economic Review*, 95, 724–738.

De Villiers, D. and Phiri, A. (2022) “Towards resolving the purchasing power parity (PPP) ‘Puzzle’ in newly industrialized countries (NIC’s)”, *The Journal of International Trade & Economic Development*, 31(2), 161-180.

Dornbusch, R. (1976) “Expectations and Exchange Rate Dynamics”, *The Journal of Political Economy*, 84(4), 1161–1176.

Elberg, A. (2016) “Sticky prices and deviations from the Law of One Price: Evidence from Mexican micro-price data”, *Journal of International Economics*, 98, 191–203.

Feenstra, R. C., Inklaar, R. and Timmer, M. P. (2015) “The Next Generation of the Penn World Table”, *American Economic Review*, 105(8), 3150-3182.

Huang, C. H. and Yang, C. Y. (2015) “European exchange rate regimes and purchasing power parity: An empirical study on eleven eurozone countries”, *International Review of Economics and Finance*, 35, 100 – 109,

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y. (1992). “Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?” *Journal of Econometrics*, 54, 159–178.

MacDonald, R. (2007) *Exchange Rate Economics: Theories and Evidence*. Routledge (2nd edition).

Majumder, A. and Ray, R. (2020) “National and subnational purchasing power parity: a review”, *Decision*, 47(2):103–124.

Pelagatti, M. and Colombo, E. (2015) “On the empirical failure of purchasing power parity tests”, *Journal of Applied Econometrics*, 30, 904 – 923.

Robertson, R., Kumar, A. and Dutkowsky, D. H. (2014) “Weak-form and strong-form purchasing power parity between the US and Mexico: a panel cointegration investigation”, *Journal of Macroeconomics*, 42, 241–262.

Rogoff, R. (1996) "The Purchasing Power Parity Puzzle", *Journal of Economic Literature*, 34, 647–68.

Rossi, B. (2005) "Confidence Intervals for Half-life Deviations from Purchasing Power Parity", *Journal of Business and Economic Statistics*, 23(2), 432-442.

Sarno, L. and Taylor, M. P. (2002) "Purchasing power Parity and the Real Exchange Rate", *IMF Staff Papers*, 49(1), 65–105.

She, F., Zakaria, M., Khan, M. and Wen, J. (2021) "Purchasing Power Parity in Pakistan: Evidence from Fourier Unit Root Tests", *Emerging Markets Finance and Trade*, 57(9), 3835-3854.

Taylor, A. M. and Taylor, M. P. (2004) "The Purchasing Power Parity Debate" *Journal of Economic Perspectives*, 18(2), 135–158.

Taylor, M. P. (2009) "Long-Run Purchasing Power Parity and real Exchange Rates: Introduction and Overview", *Applied Economics Letters*, 16, 1–4.

Wu, T., Bahmani-Oskooee, M., Chang, T. (2018) "Revisiting purchasing power parity in G6 countries: an application of smooth time-varying cointegration approach", *Empirica*, 45, 187-196.

Zhang, Z., Bian, Z. and Zhan, M. (2022) “Is absolute purchasing power parity special for Spain”, *Empirical Economics*, 62, 513–531.

Zhang, Z. and Zou, X. (2014) “Different measures in testing absolute purchasing power parity”, *Applied Economics Letters*, 21(8), 828–831.