Long-Run Purchasing Power Parity Redux

David H. Papell and Ruxandra Prodan

University of Houston

February 16, 2018

Abstract

We test for long-run covariability between nominal exchange rate depreciation and inflation differentials to provide a new perspective on long-run Purchasing Power Parity (PPP). The method directly tests Cassel's concept of relative PPP and is more robust to departures from exact unit roots in nominal exchange rates and relative prices than standard unit root and cointegration tests for PPP. The central result of the paper is that the 90 percent confidence interval for (1) the long-run correlation coefficient is above zero and (2) the long-run linear regression coefficient contains one and, therefore, long-run PPP cannot be rejected for 9 of the 16 countries. For six of the countries, adding a third criterion that the confidence interval for the long-run linear regression coefficient have relatively narrow bands provides strong evidence of long-run PPP. We also find evidence consistent with Balassa-Samuelson effects that countries with higher per capita income have stronger real exchange effects.

JEL Classification: C32; F40

Keywords: Purchasing Power Parity; Long-Run Covariability

We thank Ulrich Müller and Mark Watson for helpful comments and discussions.

Correspondence to:

David Papell, tel. (713) 743-3807, email: dpapell@mail.uh.edu, Department of Economics, University of Houston, Houston, TX 77204-5019

Ruxandra Prodan, tel. (713) 743-3836, email: rprodan@uh.edu, Department of Economics, University of Houston, Houston, TX 77204-5019

"The general inflation which has taken place during the war has lowered this purchasing power in all countries, though in a different degree, and the rates of exchange should accordingly be expected to deviate from their old parities in proportion to the inflation in each country." Cassel (1918)

"Purchasing Power Parity (PPP) is a theory of exchange rate determination. It asserts (in the most common form) that the exchange rate change between two currencies over any period of time is determined by the change in the two countries' relative price levels." Dornbusch (1987)

1. Introduction

We use time series econometrics techniques recently developed by Müller and Watson (2017) to provide a new perspective on an old question. For the past 30 years, the overwhelming preponderance of research on Purchasing Power Parity (PPP) has involved unit root tests on real exchange rates and/or cointegration tests on nominal exchange rates and relative prices, with rejection of the unit root and/or no cointegration null hypotheses providing evidence of long-run PPP. This evidence, however, crucially depends on nominal exchange rates and relative prices having exact unit roots. We extend the literature by directly testing Cassel's proposition that exchange rate depreciation is proportional to inflation differentials over the past 143 years using a methodology that is robust to departures from the unit root model.

The evolution of research on PPP was summarized in a survey by Froot and Rogoff (1995). Stage one tests have short-run PPP as the null hypothesis. They regress the nominal exchange rate on a constant and relative prices in the two countries, with a coefficient of one on relative prices constituting evidence of PPP. By the time the survey was written, it was well-understood that these tests were fundamentally flawed because they failed to take account of potential nonstationarity of exchange rates and relative prices. Stage one tests were also, as in Frenkel (1981), conducted on a first-differenced specification between exchange rate depreciation and inflation differentials. In stage two tests, the null hypothesis is that the real exchange rate follows a random walk, with rejection of the null providing evidence of long-run PPP. While a major focus of this research was low power of unit root tests with the short spans of post-1973 flexible exchange rate data, this is not an issue with long-horizon data of a century or more. In stage three tests, the null hypothesis is that the nominal exchange rate and relative prices are cointegrated, with rejection of the null constituting evidence of a weaker form of long-run PPP. Froot and Rogoff conclude that stage three tests had not provided any insights beyond those of the stage two tests. For the past 20 years, almost all research on PPP has been conducted in the framework of stage two tests.[1]

Unit root tests on long-horizon real exchange rates have not produced strong evidence of long-run PPP. Taylor (2002) constructs U.S. dollar real exchange rates for 16 developed countries with about 120 years of data and rejects the unit root null at the 5 percent level for 6 of the 16 rates using ADF tests and 10 of the 16 rates using DF-GLS tests from Elliott, Rothenberg, and Stock (1996). Lopez, Murray, and Papell (2005), using more appropriate lag selection methods and a slightly extended version of Taylor's data, reject the unit root null at the 5 percent level for 8 of the 16 countries with ADF tests and 9 of the 16 countries with DF-GLS tests. While Papell and Prodan (2006) report additional rejections using tests for a unit root in the presence of structural change where the null hypothesis is a unit root without structural change, these tests have been shown to be oversized if the null hypothesis is a unit root with structural change. Panel unit root tests, which exploit cross-section variability and are useful for increasing power with post-Bretton Woods data, are not helpful with long-horizon data since both ADF and DF-GLS tests have sufficient power with over 100 years of annual observations.[2]

The application of unit root tests to long-run PPP raises both conceptual and statistical issues. Cassel's concept of PPP, that exchange rate depreciation should be proportional to inflation differentials, is called "relative" PPP. "Absolute" PPP, in contrast, requires that exchange rate and relative price movements offset each other, so that the real exchange rate is constant.[3] While rejection of the unit root null in real exchange rates in favor of a level stationary alternative can, in principle, provide evidence of long-run mean reversion and, therefore, long-run PPP, the only available data on national price levels are price indexes. Since these indexes are set to 100 for an arbitrarily chosen base year, the fact that one country's price level is higher, lower, or equal to another country's in any given year provides no information on actual relative prices. While the motivation for the stage two tests comes from testing absolute PPP, they are actually indirect tests of relative PPP.

[1] While the literature is too voluminous to reference here, surveys are provided by Breuer (1994) and Taylor and Taylor (2004).

[2] Engel (2000) argues that rejection of the unit root null by ADF and/or DF-GLS tests does not constitute evidence of stationarity because any stationary representation of a time series can be arbitrarily well-represented by a nonstationary representation, and explores the practical implications of this result for studies of PPP.

[3] While Cassel also wrote about absolute PPP, the best-known application of his work, determining the amount of overvaluation of the pound when the gold standard was restored at the prewar mint parity, involved relative PPP. This is described by Officer (1976).

A statistical problem with stage two and three tests is that both nominal exchange rates and relative prices are assumed to have unit roots. Tests for a unit root in real exchange rates are tests for cointegration between nominal exchange rates and relative prices with a restriction that the coefficient on relative prices in the cointegrating vector equals unity. In order for these tests to be valid, the variables need to have exact unit roots because, as shown by Elliott (1998), very small deviations from a unit root can cause very large size distortions in cointegration tests. The "knowledge" that nominal exchange rates and relative prices have unit roots, however, comes from failure to reject the unit root null for these variables using tests that have low power to distinguish between unit roots and close alternatives. As emphasized by Stock and Watson (2017), evidence of cointegration is extremely fragile to departures from exact unit roots.

Müller and Watson (2017) develop a methodology to measure "long-run covariability" between two time series. They show how to calculate point estimates and construct confidence intervals for long-run correlation coefficients, linear regression coefficients, and standard deviations of regression errors that are robust to stationary variables integrated of order zero, or I(0), unit root series integrated of order one, or I(1), near unit root and fractionally integrated series. Long run covariability tests Cassel's concept of relative PPP directly by calculating long-run linear regression coefficients between exchange rate depreciation and inflation differentials and, therefore, provides a more natural interpretation of long-run PPP than unit root tests. We call these methods stage four tests for long-run relative PPP, as they combine having PPP as the null hypothesis from the stage one tests with the long-run focus from the stage two and three tests. They can be interpreted as stage one tests of relative PPP using modern statistical methodology because long-run PPP implies that the long-run linear regression coefficient between growth rates of nominal exchange rates and relative prices should equal unity. It is important to emphasize that, in contrast to stage two and three tests, these methods do not require nominal exchange rates and relative prices to be I(1).

We use nominal U.S. dollar exchange rates for 16 countries and Consumer Price Indexes (CPI) for 17 countries from 1870 to 2013 from the Jorda, Schularick, and Taylor (2017) Macrohistory Database. The central result of the paper is that, with the "long run" for our benchmark specification consisting of six periods of 24 years, long-run PPP cannot be rejected for nine of the 16 countries. This result is based on (1) the 90 percent confidence interval for the long-run correlation coefficients do not contain zero and (2) the 90 percent confidence interval for the long-run linear regression coefficients contain one. When we add a third criterion that the confidence intervals for the long-run regression coefficients should be relatively narrow, we find strong evidence of long-run relative PPP for six of the countries. The results are similar if the "long-run" consists of nine periods of 16 years or three periods of 48 years.

While an important theme of the paper is that long-run covariability provides a better method for testing relative PPP than existent methods, we conduct Stage 2 tests in order to provide perspective on our results. Using ADF tests, the unit root null can be rejected against a level stationary alternative for 9 of the 16 real exchange rates at the 5 percent level. There is considerable overlap between the results, as long-run PPP cannot be rejected and the unit root null can be rejected for 6 of the 9 countries.

The leading explanations for non-constant long-run real exchange rates come from various versions of the Balassa-Samuelson model, where the long-run real exchange rate is a function of the relative productivity of traded to nontraded goods in the two countries. Since sectoral productivity data are not available over long spans of data, we follow Bordo et al. (2017) and use the income per capita differential between the two countries as a proxy for the for the traded goods productivity differential and test for long-run covariability between the income differential and the real exchange rate. Both the long-run correlation and linear regression coefficients are in accord with the Balassa-Samuelson model for 14 of the 16 countries, although the estimates are less precise than the long-run PPP tests.

II. Testing for Purchasing Power Parity

Purchasing Power Parity (PPP) is a term that encompasses a number of different concepts. Absolute PPP is a generalization of the Law of One Price (LOP), which states that an individual good should have identical prices expressed in a common currency across countries, to national price levels. If the LOP holds for all goods, then it must hold for any identical basket of goods. Relative PPP is that changes in relative price levels should be offset by changes in the exchange rate. Both concepts of PPP can be analyzed in either the short run or the long run.[4]

[4] There has been a proliferation of research on the LOP in recent years, much of it motivated by the availability of data on individual goods. Examples include Crucini, Telmer, and Zachariadis (2005) and Cavallo, Neiman, and Rigobon (2015). We do not consider the LOP in this paper.

Stage One Tests for Short-Run PPP

Stage one tests are linear regressions that treat PPP as the null hypothesis. Stage one tests for absolute PPP regress the exchange rate on a constant and relative prices,

$$e_t = \alpha + B(p_t - p_t^*) + \varepsilon_t , \qquad (1)$$

where e_t is the log of the exchange rate (defined as the domestic currency price of one unit of foreign exchange so that an increase in e is a depreciation of the exchange rate), p_t and p_t^* are the logs of the domestic and foreign price levels, and ε_t is an error term. Absolute PPP holds if $\alpha = 0$ and $B = 1$. Stage one tests for relative PPP regress the change in the exchange rate on the change in relative prices,

$$\Delta e_t = B\,\Delta(p_t - p_t^*) + \varepsilon_t , \qquad (2)$$

where Δe_t is the rate of depreciation and $\Delta(p_t - p_t^*)$ is domestic inflation relative to foreign inflation. Relative PPP holds if $B = 1$. Frenkel (1981) is the best-known application of stage 1 tests. He found evidence of absolute PPP in the 1920s but did not find evidence of either absolute or relative PPP in the 1970s.

Stage Two and Three Tests for Long-Run PPP

Stage two tests focus on the real exchange rate, $q_t = e_t + p_t^* - p_t$, where an increase in q is a depreciation of the real exchange rate of the domestic country. From Equation (1), absolute PPP holds if the real exchange rate (in levels) equals one so that the real exchange rate q_t (in logs) equals zero. National price levels, however, are constructed to equal 100 in some base year, which does not imply that the two countries' prices are equal in that year. This has two implications. First, PPP implies that the real exchange rate is a constant, but not necessarily one (in levels) or zero (in logs). Second, all tests of PPP using national price levels are tests of relative PPP because a constant real exchange rate does not mean that an identical basket of goods will cost the same in the two countries.

When the Bretton Woods system of fixed exchange rates collapsed and the flexible exchange rate system started in March 1973, many economists believed that, based on the experience with floating exchange rates in the 1920s, exchange rates would move in response to inflation differentials and short-run PPP would hold. It quickly became apparent, however, that real exchange rates were not constant and short-run PPP did not hold. While most theories designed to explain real exchange rate movements, notably Dornbusch (1976), imposed long-run PPP as an equilibrium condition, this could not be tested with only a few years of flexible exchange rates.

As flexible exchange rates continued into the 1980s, the combination of more data and the development of unit root tests led to stage two tests. In these tests, the null hypothesis is that the real exchange rate contains a unit root, so there is no long-run mean, and the alternative hypothesis is that the real exchange rate is level stationary, so there is long-run convergence to a constant mean. Rejection of the unit root null in favor of the level stationary alternative provides evidence of long-run PPP. The most commonly used unit root tests are Augmented-Dickey Fuller (ADF) tests, which are conducted as follows,

$$q_t = \mu + \gamma q_{t-1} + \sum_{i=1}^{k} c_i \Delta q_{t-i} + \varepsilon_t \qquad (3)$$

where the unit root null is rejected if γ is significantly less than one. Equivalently, the tests can be run by subtracting q_{t-1} from both sides of the equation,

$$\Delta q_t = \mu + (\gamma - 1)\, q_{t-1} + \sum_{i=1}^{k} c_i \Delta q_{t-i} + \varepsilon_t \qquad (4)$$

where the unit root null is rejected if γ-1 is significantly less than zero.

The most important problem with the application of stage two tests to post-Bretton Woods flexible exchange rates is the low power of ADF tests. Froot and Rogoff (1995) show that, if the half-life of PPP deviations, the time that it takes for a shock to the real exchange rate from long-run PPP to dissipate by 50 percent, is three years, it would take 72 years of data to have good power to reject the unit root null at the 5 percent level. One response to this problem was to use longer data sets that combine fixed and flexible exchange rate periods. Examples include Frankel (1986), Lothian and Taylor (1996), and Taylor (2002). Another response was to use panel unit root tests that combine cross-section and time series variability. Examples include Abuaf and Jorion (1990), Frankel and Rose (1996), and Papell (1997).

Stage three tests involve cointegration between either (1) nominal exchange rates and relative prices or (2) nominal exchange rates, domestic prices, and foreign prices. Cointegration occurs if two or more variables have unit roots or (I(1)) but a linear combination of the variables is stationary (I(0)). The most commonly used method to test for cointegration is the Engle-Granger (EG) test. In the context of testing for PPP, first estimate the bivariate cointegrating vector,

$$e_t = \alpha + B(p_t - p_t^*) + z_t\,, \qquad (5)$$

Next, run an ADF test on the residuals z_t. Evidence of cointegration is found if the null hypothesis of a unit root can be rejected in favor of the alternate hypothesis of stationarity.[5] Stage two tests are equivalent to stage three tests plus an additional restriction that B = 1.[6] While more "evidence" of PPP was often found with stage three tests than with stage two tests, it is difficult to see what is meant by "parity" without the restriction that B = 1, and the vast majority of tests for long-run PPP are unit root tests.

While the logic for stage two tests comes from absolute PPP, the tests are actually of relative PPP. This is most easily seen from Equation (3). If the real exchange rate is stationary ($\gamma < 1$), there is mean reversion of the real exchange rate and long-run PPP holds.[7] If the real exchange rate has a unit root ($\gamma = 1$), there is no mean reversion of the real exchange rate and long-run PPP does not hold. If the price levels for the two countries consisted of identical baskets of goods, this would be a test of long-run absolute PPP. But they do not, as the price levels are indexes set to the same number at the same date. While stage two tests look like tests for absolute PPP, they are actually tests for relative PPP.

Issues with Stage Two Tests for Long-Run PPP

There is an important issue with stage two tests for long-run PPP that has received very little attention. Unit root tests for long-run PPP are, as described above, stage three (cointegration) tests with additional restrictions. In order for two or more variables to be cointegrated, the variables must be integrated. For example, if one variable is I(1) and the other is I(0), every linear combination is I(1) and they cannot be cointegrated. Conversely, if both variables are I(0), every linear combination will be I(0) but they are not cointegrated. The first step in EG tests for cointegration is, therefore, to test for integration of the individual variables. In the context of testing for long-run PPP, this usually means conducting ADF and/or DF-GLS tests on relative (or individual) prices. The individual variables are taken to be I(1) if the unit root null cannot be rejected.

The problem with this methodology is that, as shown by Elliott (1998) and described less technically by Stock and Watson (2017), tests for cointegration are very sensitive to small departures from the unit root assumption. Elliott (1998) shows that tests with a nominal size of 5

[5] The ADF test is run without a constant because it is already included in Equation (5).
[6] In the trivariate version of cointegration that allows different coefficients on domestic and foreign prices, stage two tests impose a coefficient of 1 on domestic prices and a coefficient of -1 on foreign prices.
[7] The coefficient γ also has to be greater than -1, but that is never an issue in practice.

percent can be very oversized with small differences between the true largest autoregressive root and one, especially when the correlation between innovations in the error and the regressor are large. As described by Stock and Watson (2017), "this problem arises for deviations from a unit root that are too small to be detected with high probability, even in arbitrarily large samples. As a result, standard methods of inference developed for cointegration models are not robust to effectively undetectable departures from the model, making such inference unreliable."[8]

What are the implications of these results for stage two tests of long-run PPP? If we knew for sure that nominal exchange rates and relative prices had exact unit roots, we could test for long-run PPP by using tests for cointegration with restrictions, i.e. unit root tests on real exchange rates. But we can never know this for sure. As expressed by Stock and Watson (2017), "there are clearly low-frequency comovements in the data, and macroeconometricians need a set of tools for quantifying those comovements that does not hinge on adapting a particular model, such as a unit root model, for the underlying trends."[9]

Long-Run Covariability

Müller and Watson (2017) develop tests for long-run covariability between two time series that do not depend on a particular model. There is a large amount of technical material in the paper, and we will only sketch the ideas here. The methods use low-frequency averages of the data to measure long-run variability and covariability. They are defined in terms of population second moments of low frequency trends that are similar to low-pass filtered data. In order to handle serial correlation, they use projections on low-frequency periodic (cosine) functions. They derive confidence intervals for long-run correlation coefficients, linear regression coefficients, and standard deviations of regression errors that are valid over a wide range of persistence parameters, including I(0), I(1), near unit roots, and fractionally integrated variables and are "nearly optimal" in the sense of having close to the shortest expected length. They focus on two empirical examples: GDP and consumption and short-and long-term nominal interest rates.

These methods are ideally suited for investigating long-run relative PPP. The first test statistic is the long-run correlation coefficient between exchange rate depreciation and relative inflation. If the confidence interval of the correlation coefficient ρ lies above zero, then countries with higher inflation than the U.S. will have depreciating currencies, consistent with long-run

[8] Stock and Watson (2017), p. 81.
[9] Stock and Watson (2017), p. 83.

relative PPP. The second test statistic is the long-run linear regression coefficient B between exchange rate depreciation and relative price inflation. If the confidence interval around B contains one and does not contain zero, the evidence is consistent with long-run relative PPP. This is a long-run version of the stage one test for relative PPP in Equation (2). In both cases, relative PPP is the null hypothesis and the failure to reject the null that B = 1 provides evidence of relative PPP.

There are two concepts of "long run" that are used for the analysis. The first is statistical. The data are averaged over periods consisting of the number of observations divided by $x/2$, where x is chosen by the researcher. In Müller and Watson (2017), the baseline value for x is 12 and the span of the data is about 65 years, so the periods are large relative to the sample size. The second is economic. Our baseline value for x is also 12, so a period with 143 observations is 24 years. This should be long enough for adjustments to PPP to be completed if long-run PPP holds. We also report results with x = 18, where the long run is 16 years and x = 6, where the long run is 48 years.[10]

In addition to their major empirical examples, Müller and Watson (2017) consider a number of other long-run relations, including nominal exchange rate depreciation and relative inflation for the United States and the United Kingdom from 1971-2015. They do not find evidence of long-run relative PPP, as the 90 percent confidence interval for ρ contains zero and the 90 percent confidence interval for B is too large to be informative. The "long run" for this example is approximately 7 years, which they speculate may be too short for long-run adjustments to have occurred. This is consistent with research on median-unbiased half-lives of PPP deviations, such as Murray and Papell (2002), which generally report results between 3 and 5 years with very wide confidence intervals.

III. Empirical Results on Long-Run Purchasing Power Parity

The research in this paper has been greatly enhanced by the availability of the Jorda, Schularick, and Taylor (2017) Macrohistory Database. We use nominal U.S. dollar exchange rates for 16 countries and construct relative prices from Consumer Price Indexes (CPI) for the same countries (relative to the United States) from 1870 to 2013.[11] The countries are listed in Table 1.

[10] Müller and Watson (2017) use q to denote the length of the period. We use x because q is the standard notation for the real exchange rate.
[11] The data is available at www.macrohistory.net/data/

Exchange rate depreciation and relative inflation are calculated by taking first-differences of the logs of exchange rates and relative prices.

The most general concept of long-run relative PPP is provided in Figure 1, which depicts the average annual nominal exchange rate depreciation and relative price inflation for 15 of the 16 countries vis-à-vis the United States between 1870 and 2013.[12] The averages are clustered around the upward-sloping 45 degree line, which illustrates the well-known result that high inflation countries have long-run depreciating nominal exchange rates.

Graphical Intuition

We start by providing some graphical intuition for our subsequent statistical results. Figure 2 depicts the average annual nominal exchange rate depreciation and relative price inflation for 6 non-overlapping periods of 24 years for each of the 16 countries vis-à-vis the United States. This corresponds to the "long-run" with 143 annual observations and x = 12. Points on or close to the 45 degree line are consistent with long-run relative PPP because the average rate of depreciation equals the average inflation differential.

The graphical results show a wide variety of experiences. France is an example of a country where the points are clustered along the 45 degree line, indicating (visual) consistency with long-run relative PPP. For Belgium, some, but not all, of the points are close to the 45 degree line and, for Sweden, there is no visual evidence of PPP. Overall, the visual evidence is mixed. Germany requires further explanation. The figure depicts five, instead of six, points, which are fairly close to the 45 degree line. The missing sixth point, for 1919 – 1942 which includes the German hyperinflation, has an average annual inflation differential of 111.48 percent and an average annual exchange rate depreciation of 111.46 percent with the United States and renders the other five points unreadable.;

Long-Run Covariability Tests

The central results of the paper are in Table 1, which presents the long-run correlation coefficients ρ and the long-run linear regression coefficients B for the 16 countries with x = 12. The point estimate of ρ is positive for all 16 countries, indicating that high inflation countries relative to the U.S. have long-run depreciating exchange rates. The correlations are very spread out, with a low of 0.04 and a high of 0.99, with 9 of the 16 correlations above 0.5. The 90 percent

[12] Germany is excluded from Figure 1, although not from the subsequent analysis, because the hyperinflation period raises its average inflation and depreciation by so much as to make the figure unreadable.

confidence interval is bounded above zero for 10 of the 16 countries. The point estimate of B is also positive for all 16 countries, with a low of 0.10 and a high of 1.21. The 90 and 95 percent confidence intervals contain one for 8 of the 16 countries.

As with the stage 1 tests, evidence of PPP comes from failure to reject the null hypothesis, in this case long-run PPP. For nine of the 16 countries, Belgium, Finland, France, Germany, Italy, Japan, Portugal, Switzerland, and the United Kingdom, the 90 percent confidence interval around ρ lies above zero and the 90 percent confidence interval around B contains one and does not contain zero. For six of these nine countries, Finland, France, Germany, Italy, Japan, and Portugal, the point estimate of ρ is close to one and the confidence interval around B is tight, providing strong evidence of long-run PPP. We discuss these six countries below.

Finland – The point estimate of ρ is 0.96, with a 90 percent confidence interval of 0.90 – 0.98. The point estimate of B is 1.21, the highest among the 16 countries, with a 90 percent confidence interval of 1.03 - 1.39 and a 95 percent confidence interval of 0.98 – 1.44. This is the only country for which the 95 percent confidence interval contains one but the 90 percent confidence interval does not.

France - The point estimate of ρ is 0.97, with a 90 percent confidence interval of 0.93 – 0.98. The point estimate of B is 1.20 with a 90 percent confidence interval of 0.96 – 1.39.

Germany - The point estimate of ρ is 0.99, with a 90 percent confidence interval of 0.98 – 0.99. The point estimate of B is 1.00 with a 90 percent confidence interval of 0.96 – 1.03. Among the six countries, Germany provides the strongest evidence of long-run PPP, with the point estimate of ρ the highest and the point estimate of B the closest to one. This is consistent with previous research that the strongest evidence of PPP comes from hyperinflation episodes.

Italy - The point estimate of ρ is 0.96, with a 90 percent confidence interval of 0.90 – 0.98. The point estimate of B is 1.13 with a 90 percent confidence interval of 0.96 – 1.31.

Japan - The point estimate of ρ is 0.97, with a 90 percent confidence interval of 0.90 – 0.99. The point estimate of B is 0.98 with a 90 percent confidence interval of 0.85 – 1.20.

Portugal - The point estimate of ρ is 0.96, with a 90 percent confidence interval of 0.89 – 0.98. The point estimate of B is 1.14 with a 90 percent confidence interval of 0.95 – 1.32.

For three of the nine countries, Belgium, Switzerland, and the United Kingdom, while the null hypothesis of long-run PPP cannot be rejected because the 90 percent confidence interval around ρ lies above zero and the 90 percent confidence interval around B contains one and does not contain zero, the combination of lower point estimates for ρ and wider confidence intervals around B leads us to call this marginal evidence of long-run PPP. We discuss these three countries below.

Belgium - The point estimate of ρ is 0.64, with a 90 percent confidence interval of 0.12 – 0.89. The point estimate of B is 0.66 with a 90 percent confidence interval of 0.17 – 1.13.

Switzerland - The point estimate of ρ is 0.45, with a 90 percent confidence interval of 0.01 – 0.78. The point estimate of B is 0.62 with a 90 percent confidence interval of 0.08 – 1.23.

United Kingdom - The point estimate of ρ is 0.65, with a 90 percent confidence interval of 0.23 – 0.87. The point estimate of B is 0.92 with a 90 percent confidence interval of 0.39 – 1.48.

For the seven remaining countries, Australia, Canada, Denmark, Spain, Netherlands, Norway, and Sweden, the long run covariability tests do not provide evidence of long-run PPP. For each country, either the confidence interval around ρ contains zero and/or the confidence interval around B does not contain one.

Australia - The point estimate of ρ is 0.38, with a 90 percent confidence interval of -0.06 – 0.72. The point estimate of B is 0.45 with a 90 percent confidence interval of -0.07 – 1.00.

Canada - The point estimate of ρ is 0.39, with a 90 percent confidence interval of -0.08 – 0.80. The point estimate of B is 0.41 with a 90 percent confidence interval of -0.06 – 0.95.

Denmark - The point estimate of ρ is 0.04, with a 90 percent confidence interval of -0.38 – 0.48. The point estimate of B is 0.10 with a 90 percent confidence interval of -0.64 – 0.83.

Netherlands - The point estimate of ρ is 0.38, with a 90 percent confidence interval of -0.06 – 0.72. The point estimate of B is 0.73 with a 90 percent confidence interval of -0.09 – 1.51.

Norway - The point estimate of ρ is 0.30, with a 90 percent confidence interval of -0.16 – 0.66. The point estimate of B is 0.50 with a 90 percent confidence interval of -0.31 – 1.31.

Spain - The point estimate of ρ is 0.50, with a 90 percent confidence interval of 0.01 – 0.89. The point estimate of B is 0.47 with a 90 percent confidence interval of 0.01 – 0.81.

Sweden - The point estimate of ρ is 0.08, with a 90 percent confidence interval of -0.33 – 0.54. The point estimate of B is 0.14 with a 90 percent confidence interval of -0.39 – 0.67.

These results are robust to different definitions of the “long run”. The results for x = 18 and x = 6 with 90 percent confidence intervals are also reported in Table 1. For x = 18, where the long run is 16 years, strong evidence of PPP is found for the same six countries, Finland, France, Germany, Italy, Japan, and Portugal, as with x = 12, although marginal evidence of PPP is only found for Spain. For x = 6, where the long run is 48 years, strong evidence of PPP is found for Belgium, France, Germany, Italy, Japan, and Portugal, with marginal evidence for Finland and Spain.

The graphical intuition and econometric results are positively, although not perfectly, correlated. Among the six countries for which we find strong evidence of PPP with x = 12, France, Germany, Japan, and Portugal have graphical results that cluster around the 45 degree line while, for Finland and Italy, four of the six sub-periods cluster around the 45 degree line while the other two sub-periods deviate in opposite directions. Among the other 10 countries, none of the sub-periods cluster particularly closely around the 45 degree line. While the statistical results differentiate three countries with marginal evidence of PPP; Belgium, Switzerland, and the United Kingdom, from the seven others, this is not obvious from the graphical intuition. Overall, the econometric methods provide a much sharper differentiation among the countries than the graphical intuition.

Comparison with Stage Two Tests

The standard method for investigating PPP is to conduct unit root tests on real exchange rates. While the motivation for the paper comes from the argument that long-run covariability tests are preferable to stage two tests because (1) they are direct tests of the hypothesis of relative PPP and (2) they are robust to departures from unit roots in nominal exchange rates and relative prices, we present stage two tests for comparison with previous research.

The results are reported in Table 2. The unit root null is rejected in favor of the alternative of level stationarity with ADF tests for 9 of the 16 countries, approximately the same as in Taylor (2002) and Lopez, Murray, and Papell (2005). Out of the six countries for which we find strong evidence of PPP using long-run covariability tests, the unit root null can be rejected a the 5 percent level for four countries and at the 10 percent level for one additional country. Among the three countries for which we find marginal evidence of PPP using long-run covariability tests, the unit root null can be rejected a the 5 percent level for two countries. While the results of the long-run covariability tests are positively correlated with the results of the stage two tests, they are not identical. The evidence for long-run PPP is stronger for Japan and Switzerland with the long-run covariability tests, and is stronger for Canada, Netherlands, and Sweden with the stage two tests.[13]

III. The Balassa-Samuelson Model

The leading explanation for long-run real exchange rate movements comes from the Balassa-Samuelson model. In the "standard" version of the model, long-run PPP holds for traded goods and the long-run real exchange rate depends on the relative productivity of traded to nontraded goods in the two countries. There is a vast literature on the Balassa-Samuelson model, including "modern" versions where long-run PPP holds for neither traded nor nontraded goods.

While sectoral productivity data are available for the post-Bretton Woods period, they are not available for earlier periods. We follow the usual practice, as in Bordo et al. (2017), and use the per capita income differential between the two countries as a proxy for the traded goods productivity differential.[14] In that case, the Balassa-Samuelson effect is identified by a positive coefficient on the per capita income differential in a real exchange rate regression.

$$q_t = \alpha + B(y_t^* - y_t) + \varepsilon_t , \tag{1}$$

where $q_t = e_t + p_t^* - p_t$ is the log of the real exchange rate for the domestic country, y_t is the log of domestic per capita income, and y_t^* is the log of foreign per capita income. If foreign per capita income is higher than domestic per capita income, the real exchange rate of the foreign country will be appreciated relative to its PPP value.

Tests of the Balassa-Samuelson model are subject to even more severe problems than Stage 2 tests of long-run PPP. In order for the point estimates and standard errors of B in Equation (3) to be valid, both the real exchange rate and relative per capita income need to be stationary. The evidence for long-run real exchange rate stationarity, however, is clearly mixed, both with Stage 2 tests and with the long-run covariability tests presented above. There is also no consensus in the literature on whether per capita income is stationary or has a unit root.[15] We avoid these problems by using long-run covariability tests for which the real exchange rate and relative per capita income can be either I(0) and/or I(1).

---

[13] We also ran DF-GLS tests. The rejections of the unit root null were weaker than with the ADF tests, and the results were very sensitive to lag lengths.

[14] Balassa (1964), Samuelson (1964), and Lothian and Taylor (2008) also use this proxy.

[15] Bordo et al. (2017) run panel unit root tests on long-term real exchange rates and relative per capita income. They either use Levin, Lin, and Chu (2002) tests, which are oversized in mixed panels, or Im, Pesaran, and Shin (2003) and Peseran (2003) tests, for which the alternative hypothesis is that at least one member of the panel is stationary.

Long-Run Covariability Tests

Table 3 presents the long-run linear regression coefficients B for the 16 countries with x = 12. The point estimate of B, which corresponds to the standard regression tests of the Balassa-Samuelson model in Equation (3), is positive for 14 of the 16 countries, indicating that countries with higher per capita income have appreciated real exchange rates. Belgium and Switzerland are the only countries for which the results are not consistent with the model. The estimates, however, are not as precise as those for the long-rum PPP tests. While the 67% confidence interval is bounded above zero for 9 of the 14 countries, Australia, Denmark, Finland, France, Germany, Italy, Japan, Norway, and Spain, the 90% confidence interval is only bounded above zero for Germany, Italy, and Japan and the 95% confidence interval is only bounded above zero for Italy, and Japan.

The long-run correlation coefficients ρ for the 16 countries with x = 12 are also presented in Table 3. The results are very similar to those for the long-run linear regression coefficients B, with the point estimate of ρ positive for the same 14 countries, strengthening the finding that countries with higher per capita income have appreciated real exchange rates. The 67%, 90%, and 95% confidence intervals are bounded above zero for 7, 3, and 1 of the 14 countries. While the confidence intervals for the long-run regression coefficient B are wider for the tests of the Balassa-Samuelson model in Table 3 than for the tests of long-run PPP in Table 1, the precision is not directly comparable because the tests for long-run PPP use nominal exchange rate and price data directly while the tests for the Balassa-Samuelson model proxy sectoral productivity differentials with per capita income differentials.

The results for x = 18 and x = 6 with 90 percent confidence intervals, reported in Table 3, are similar to those for x = 12. For both x = 18, where the long run is 16 years, and x = 6, where the long run is 48 years, the point estimates for B and ρ are positive for 13 of the 16 countries. In addition to Belgium and Switzerland, the results are not consistent with the model for the United Kingdom. As with x = 12, the 90 percent confidence intervals are wide, being bounded above zero for the minority of countries in all cases.

Bordo et al. (2017) find evidence that is generally consistent with the Balassa-Samuelson model for their long-run sample. Since they use panel methods, it is not possible to make a country-by-country comparison with their results. But as with our long-run PPP results, such comparisons would be of second-order importance. The main takeaway of the results in this section is that, in contrast with conventional tests, we have provided evidence consistent with the Balassa-Samuelson model that is robust to departures from stationarity of both real exchange rates and per capita income differentials.

IV. Conclusions

Purchasing Power Parity has been an important topic in international economics for the past century. Over the past 20 years, evidence of PPP has become virtually synonymous with rejection of a unit root in real exchange rates in favor of level stationarity. We propose an alternative methodology based on the long-run covariability tests of Müller and Watson (2017). In line with the Froot and Rogoff (2005) nomenclature, we call these stage four tests because they combine having PPP as the null hypothesis from stage one tests with the long-run focus from stage two and three tests. They provide an intuitive description of PPP and, in contrast to unit root tests, are robust to small departures from non-stationarity in nominal exchange rates and relative prices.

We test for long-run relative PPP using 143 years of annual nominal exchange rate and relative price data from 1870 to 2013 for 16 countries vis-vis the United States from the Jorda, Schularick, and Taylor (2017) Macrohistory Database. We find evidence of PPP based on two tests: the confidence intervals for the long-run correlation coefficients do not contain zero and the confidence intervals for the long-run linear regression coefficients contain one. With a long-run of six periods of 24 years and 90 percent confidence intervals, long-rum PPP cannot be rejected for nine of the 16 countries. Adding a third, albeit less precise, criterion that that the confidence intervals for the long-run regression coefficients should be relatively narrow, we find strong evidence of long-run relative PPP for six of the countries. The results are similar if the "long-run" consists of nine periods of 16 years or three periods of 48 years or if 67 percent or 95 percent confidence intervals are used. We also find evidence in accord with the Balassa-Samuelson model of real exchange rate determination for 14 of the 16 countries, although the estimates are less precise than the long-run PPP tests.

There is considerable accord between our results and conventional methods. Using Augmented-Dickey-Fuller tests, the unit root null can be rejected against a level stationary alternative for 9 of the 16 real exchange rates at the 5 percent level. For six countries, long-run PPP cannot be rejected using long-run covariability tests and the unit root null can be rejected using ADF tests. The main point of the paper, however, is not whether our evidence of long-run

PPP is stronger, weaker, similar, or different than the results from conventional tests. It is that long-run covariability between nominal exchange rate depreciation and relative inflation provides a superior methodology for testing long-run PPP than either unit root teats of real exchange rates or cointegration tests between nominal exchange rates and relative prices. Our results are both intuitive and robust to departures from unit roots in nominal exchange rates and relative prices that are too small to be detected.

References

Abuaf, Niso, and Philippe Jorion, (1990). "Purchasing Power Parity in the Long Run." *The Journal of Finance*, 45, 157-74.

Bordo, Michael, Ehsan Choudhri, Giorgio Fazio, and Ronald MacDonald, (2017). "The Real Exchange Rate in the Long Run: Balassa-Samuelson Effects Reconsidered," *Journal of International Money and Finance*, 75, 69-92.

Balassa, Bela, (1964). "The Purchasing Parity Power Doctrine: a Reappraisal." *The Journal of Political Economy*, 72, 584-96.

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

Cassel, Gustav, (1918). "Abnormal Deviations in International Exchanges." *The Economic Journal*, 28, 413-15.

Cavallo, Alberto, Brent Neiman, and Roberto Rigobon, (2015). "The Price Impact of Joining a Currency Union: Evidence from Latvia, *International Monetary Fund Economic Review*, 63, 281-297

Crucini, Mario, Chris Telmer, and Marios Zachariadis, (2005). "Understanding European Real Exchange Rates," *American Economic Review*, 95, 724-738

Dornbusch, Rudiger, (1976). "Expectations and Exchange Rate Dynamics," *Journal of Political Economy*, 84, 1161-1176.

Dornbusch, Rudiger, (1987). "Purchasing Power Parity," in Eatwell, John, Murray Milgate, and Peter Newman, eds., in *The New Palgrave: A Dictionary of Economics*, 1975-1085.

Elliott, Graham, (1998). "On the Robustness of Cointegration Methods when Regressors Almost Have Unit Roots," *Econometrica*, 66,

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

Engel, Charles, (2000). "Long-Run PPP may not hold After All," *Journal of International Economics*, 57, 243-273.

Frankel, Jeffrey, (1986). "International Capital Mobility and and Crowding Out in the U.S. Economy: Imperfect Integration of Financial Markets or Goods Markets," in Hafer, R.W., ed., *How Open is the U.S. Economy*, (Lexington Books, Lexington).

Frankel, Jeffrey, and Andrew Rose, (1996). “A Panel Project on Purchasing Power Parity: Mean Reversion Within and Between Countries.” *Journal of International Economics* 40, 209-224.

Frenkel, Jacob, (1981). The Collapse of Purchasing Power Parity During the 1970s,” *European Economic Review*, 16, 145-165.

Froot, Kenneth and Kenneth Rogoff, (1995). "Perspectives on PPP and Long-Run Real Exchange Rates," in Gene Grossman and Kenneth Rogoff, eds., *The Handbook of International Economics*, Vol. 3, 1647-1688.

Im, Kyung So, Hashem Pesaran, and Yongcheol Shin, (2003). “Testing for Unit Roots in Heterogeneous Panels.” *Journal of Econometrics* 115, 53-74.

Jorda, Oscar, Moritz Schularick, and Alan Taylor, (2017). In Eichenbaum, Martin and Jonathan Parker, eds., *NBER Macroeconomics Annual 2016*, 31, 213-263

Levin, Andrew, Chien-Fu Lin, and James Chu, (2002). “Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties.” *Journal of Econometrics* 108, 1-24.

Lopez, Claude, Christian J. Murray, and David H., Papell, (2005). “State of the Art Unit Root Tests and Purchasing Power Parity.” *Journal of Money, Credit, and Banking*, 37, 361-69.

Lothian, James R., and Mark P. Taylor, (1996). “Real Exchange Rate Behavior: The Recent Float from the Perspective of the Past Two Centuries.” *Journal of Political Economy*, 104, 488-509.

Lothian, James R., and Mark P. Taylor, (2008). “Real Exchange Rates over the Past Two Centuries: How Important are the Harrod-Balassa-Samuelson Effects?” *Economic Journal*, 118, 1742-1763,

Müller, Ulrich and Mark Watson, (2017). “Long-Run Covariability,” forthcoming, *Econometrica*.

Murray, Christian J., and David H. Papell, (2002). “The Purchasing Power Parity Persistence Paradigm.” *Journal of International Economics*, 56, 1-19.

Officer, Lawrence, (1976). “The Purchasing-Power-Parity Theory of Exchange Rates: A Review Article,” *International Monetary Fund Staff Papers*, 26, 1-60.

Papell, David, (1997). “Searching for Stationarity: Purchasing Power Parity Under the Current Float.” *Journal of International Economics* 43, 313-332.

Papell, David and Ruxandra Prodan, (2006). “Additional Evidence of Long-Run Purchasing Power Parity with Restricted Structural Change,” *Journal of Money, Credit and Banking*, 38, 1329-1349.

Pesaran, Hashem, (2007). “A Simple Panel Unit Root Test in the Presence of Cross-Section Dependence,” *Journal of Applied Econometrics*, 22, 265-312

Samuelson, Paul A., (1964). “Theoretical Notes on Trade Problems.” *The Review of Economics and Statistics* 46, 145-54

Stock, James and Mark Watson, (2017). “Twenty Years of Time Series Econometrics in Ten Pictures,” *Journal of Economic Perspectives*, 31, 59-86

Taylor, Alan, (2002). “A Century of Purchasing-Power Parity.” *Review of Economics and Statistics*, 84, 139-50.

Taylor, Alan and Mark Taylor, (2004). “*The Purchasing Power Parity Debate*,” Journal of Economic Perspectives, 18, 135-158.

Table 1. Long Run Covariability Tests: Nominal Exchange Rate Depreciation and Relative CPI Inflation

a) x = 12

Country	ρ	67% CI	90% CI	95% CI	β	67% CI	90% CI	95% CI
Australia	0.384	0.070, 0.638	-0.063, 0.721	-0.150, 0.777	0.448	0.156, 0.739	-0.070, 0.998	-0.200, 1.258
Belgium	0.640	0.335, 0.750	0.124, 0.885	0.004, 0.905	0.661	0.382, 0.909	0.166, 1.126	0.073, 1.219
Canada	0.385	0.081, 0.639	-0.079, 0.800	-0.160, 0.862	0.406	0.133, 0.653	-0.059, 0.954	-0.168, 1.227
Denmark	0.036	-0.148, 0.335	-0.377, 0.477	-0.428, 0.573	0.096	-0.317, 0.509	-0.639, 0.831	-0.777, 1.014
Finland	0.961	0.931, 0.974	0.895, 0.984	0.870, 0.987	1.214	1.109, 1.308	1.027, 1.390	0.980, 1.437
France	0.971	0.954, 0.980	0.929, 0.984	0.894, 0.984	1.204	1.113, 1.306	0.961, 1.387	0.900, 1.427
Germany	0.988	0.985, 0.990	0.975, 0.990	0.970, 0.990	0.996	0.972, 1.015	0.956, 1.034	0.948, 1.042
Italy	0.962	0.927, 0.974	0.903, 0.984	0.870, 0.986	1.132	1.033, 1.232	0.955, 1.309	0.922, 1.342
Japan	0.967	0.947, 0.981	0.916, 0.988	0.834, 0.990	0.982	0.904, 1.051	0.850, 1.198	0.727, 1.260
Netherlands	0.380	0.079, 0.639	-0.063, 0.719	-0.131, 0.777	0.733	0.269, 1.197	-0.091, 1.506	-0.246, 1.712
Norway	0.301	-0.006, 0.511	-0.160, 0.664	-0.281, 0.702	0.501	0.046, 0.957	-0.309, 1.311	-0.511, 1.514
Portugal	0.961	0.929, 0.974	0.888, 0.984	0.865, 0.987	1.139	1.040, 1.239	0.952, 1.316	0.919, 1.360
Spain	0.502	0.216, 0.712	0.005, 0.885	-0.054, 0.922	0.469	0.409, 1.155	0.005, 0.807	-0.048, 1.612
Sweden	0.079	-0.098, 0.364	-0.334, 0.538	-0.385, 0.664	0.136	-0.162, 0.434	-0.394, 0.665	-0.527, 0.798
Switzerland	0.448	0.212, 0.684	0.011, 0.778	-0.041, 0.800	0.62	0.315, 0.958	0.079, 1.228	-0.023, 1.431
UK	0.651	0.438, 0.782	0.234, 0.870	0.115, 0.894	0.916	0.600, 1.232	0.389, 1.478	0.249, 1.583

b) x = 6

Country	ρ	90% CI	β	90% CI
Australia	0.547	-0.741, 1.912	0.598	-0.741, 1.912
Belgium	0.986	0.952, 0.990	0.976	0.911, 1.063
Canada	0.404	-0.212, 0.862	0.532	-0.310, 1.472
Denmark	0	-0.568, 0.762	-0.008	-1.153, 1.607
Finland	0.841	0.364, 0.984	1.089	0.571, 1.557
France	0.963	0.845, 0.989	1.038	0.817, 1.289
Germany	0.988	0.960, 0.990	0.997	0.971, 1.025
Italy	0.954	0.762, 0.986	1.029	0.736, 1.322
Finland	0.841	0.364, 0.984	1.089	0.571, 1.557
France	0.963	0.845, 0.989	1.038	0.817, 1.289
Italy	0.954	0.762, 0.986	1.029	0.736, 1.322
Japan	0.961	0.825, 0.987	1.041	0.750, 1.294
Netherlands	0.421	-0.206, 0.885	0.788	-0.324, 1.899
Norway	-0.099	-0.702, 0.604	-0.359	-1.659, 1.156
Portugal	0.961	0.834, 0.988	1.019	0.786, 1.301
Spain	0.822	0.350, 0.964	1.214	0.424, 1.860
Sweden	0.301	-0.335, 0.762	0.436	-0.563, 1.435
Switzerland	0.447	-0.210, 0.947	0.941	-0.520, 2.177
UK	0.597	-0.022, 0.947	0.630	0.062, 1.262

c) x = 18

Country	ρ	90% CI	β	90% CI
Australia	0.144	-0.131, 0.550	0.277	-0.248, 0.802
Belgium	0.418	0.004, 0.716	0.511	0.042, 0.949
Canada	0.088	-0.161, 0.462	0.246	-0.400, 0.891
Denmark	0.078	-0.301, 0.380	0.161	-0.522, 0.844
Finland	0.882	0.705, 0.924	1.206	0.920, 1.472
France	0.911	0.822, 0.964	1.215	1.022, 1.437
Germany	0.989	0.984, 0.990	1.007	0.974, 1.043
Italy	0.907	0.782, 0.961	1.139	0.928, 1.351
Finland	0.882	0.705, 0.924	1.206	0.920, 1.472
France	0.911	0.822, 0.964	1.215	1.022, 1.437
Italy	0.907	0.782, 0.961	1.139	0.928, 1.351
Japan	0.950	0.877, 0.974	0.995	0.857, 1.132
Netherlands	0.162	-0.133, 0.550	0.520	-0.439, 1.415
Norway	0.079	-0.210, 0.380	0.130	-0.434, 0.734
Portugal	0.907	0.781, 0.962	1.094	0.895, 1.293
Spain	0.408	0.003, 0.718	0.732	0.074, 1.347
Sweden	-0.006	-0.333, 0.332	-0.043	-0.539, 0.488
Switzerland	0.212	-0.100, 0.510	0.264	-0.212, 0.836
UK	0.200	-0.100, 0.639	0.260	-0.199, 0.718

Note: ρ is the long-run correlation coefficient and B is the long-run linear regression coefficient.

Table 2. Augmented-Dickey-Fuller Test for Dollar Real Exchange Rates

Countries	lags	t-statistics	p-values
Australia	0	-2.76	0.067*
Belgium	4	-2.96	0.040**
Canada	0	-2.92	0.045**
Denmark	6	-1.33	0.609
Finland	6	-2.78	0.063*
France	0	-4.91	0.000***
Germany	5	-3.71	0.004***
Italy	1	-4.98	0.000***
Japan	2	-1.45	0.552
Netherlands	1	-3.33	0.015**
Norway	6	-1.82	0.365
Portugal	1	-2.97	0.039**
Spain	7	-1.88	0.337
Sweden	2	-3.68	0.005***
Switzerland	6	-0.40	0.905
UK	8	-3.59	0.007***

Note: The lag length is selected by the general to specific method. *, **, *** denote significance at the 10%, 5%, and 1% level of significance, respectively.

Table 3. Long Run Covariability Tests: Real Exchange Rate and Per Capita Income Differential

a) x = 12

Country	ρ	67% CI	90% CI	95% CI	β	67% CI	90% CI	95% CI
Australia	0.428	.184, 0.716	-0.197, 0.834	-0.291, 0.862	0.625	0.206, 1.044	-0.171, 1.379	-0.338, 1.546
Belgium	-0.524	-0.712, -0.100	0.834, 0.291	-0.845, 0.380	-0.839	-1.239, -0.074	-1.603, 0.363	-1.749, 0.472
Canada	0.031	-0.462, 0.386	-0.604, 0.596	-0.716, 0.659	0.209	-0.466, 0.658	-0.841, 0.996	-1.029, 1.146
Denmark	0.428	0.103, 0.604	-0.061, 0.800	-0.161, 0.862	1.155	0.419, 1.974	-0.154, 3.202	-0.400, 3.611
Finland	0.269	-0.013, 0.537	-0.198, 0.762	-0.380, 0.800	0.322	0.002, 0.667	-0.269, 0.938	-0.417, 1.184
France	0.304	0.000, 0.537	-0.162, 0.834	-0.273, 0.885	0.312	0.039, 0.558	-0.153, 0.996	-0.262, 1.215
Germany	0.707	0.512, 0.841	0.197, 0.894	-0.100, 0.917	1.168	0.829, 1.508	0.074, 1.772	-0.152, 1.923
Italy	0.709	0.502, 0.862	0.197, 0.947	-0.000, 0.956	0.657	0.463, 0.869	0.180, 1.029	0.038, 1.117
Japan	0.713	0.380, 0.922	0.197, 0.964	.100, 0.964	0.840	0.390, 1.118	0.182, 1.291	0.078, 1.360
Netherlands	0.157	-0.157, 0.448	-0.365, 0.762	-0.604, 0.834	0.174	-0.219, 0.785	-0.524, 1.483	-1.091, 1.701
Norway	0.317	0.013, 0.716	-0.106, 0.800	-0.273, 0.862	0.431	0.043, 0.770	-0.296, 1.109	-0.442, 1.351
Portugal	0.209	-0.103, 0.604	-0.321, 0.800	-0.401, 0.834	0.214	-0.135, 0.750	-0.483, 1.044	-0.644, 1.152
Spain	0.319	0.041, 0.762	-0.102, 0.862	-0.269, 0.905	0.426	0.092, 0.704	-0.186, 1.038	-0.325, 1.261
Sweden	0.267	-0.006, 0.604	-0.230, 0.762	-0.302, 0.834	0.277	-0.017, 0.548	-0.266, 0.774	-0.401, 0.887
Switzerland	-0.603	-0.905, -0.379	-0.947, -0.133	-0.956, -0.027	-1.134	-1.815, -0.658	-2.223, -0.385	-2.393, 0.023
UK	0.168	-0.291, 0.477	-0.462, 0.648	-0.537, 0.667	0.166	-0.140, 0.638	-0.334, 0.832	-0.334, 0.915

Note: ρ is the long-run correlation coefficient and B is the long-run linear regression coefficient.

Country	ρ	90% CI	β	90% CI
Australia	0.377	-0.291,0.922	0.386	-0.459, 1.019
Belgium	-0.186	-0.800, 0.537	-0.187	-1.011, 0.499
Canada	0.042	-0.800, 0.762	0.112	-0.183, 1.208
Denmark	0.653	0.030, 0.922	2.407	0.421, 5.510
Finland	0.253	-0.335, 0.762	0.309	-0.758, 1.775
France	0.470	-0.158, 0.935	0.540	-0.120, 1.420
Germany	0.443	-0.380,0.885	0.726	-0.584, 2.166
Italy	0.645	0.000, 0.964	0.580	0.093 to 1.141
Japan	0.95	0.716, 0.985	1.143	0.766, 1.472
Netherlands	0.381	-0.462, 0.834	0.563	-0.777, 1.703
Norway	0.438	-0.130, 0.905	0.556	-0.256, 1.657
Portugal	0.408	-0.380, 0.885	0.403	-0.375, 1.217
Spain	0.680	0.041, 0.956	0.641	0.028, 1.281
Sweden	0.335	-0.269, 0.885	0.393	-0.344, 1.181
Switzerland	-0.643	-0.970, 0.000	-1.411	-2.464, 0.081
UK	-0.118	-0.716, 0.413	-0.109	-0.710, 0.555

18

Country	P	90% CI	β	90% CI
Australia	0.319	-0.100, 0.800	0.585	-0.166, 1.228
Belgium	-0.480	-0.834, 0.100	-0.731	-1.372, 0.216
Canada	0.028	-0.604, 0.537	0.133	-0.821, 0.801
Denmark	0.296	-0.098, 0.664	0.723	-0.178, 2.283
Finland	0.184	-0.206, 0.716	0.254	-0.319, 0.782
France	0.418	-0.008, 0.800	0.383	-0.003, 0.939
Germany	0.714	0.197, 0.902	1.314	0.279, 1.863
Italy	0.716	0.291, 0.922	0.755	0.259, 1.119
Japan	0.538	0.162, 0.922	0.752	0.163, 1.210
Netherlands	0.198	-0.380, 0.716	0.257	-0.542, 1.230
Norway	0.273	-0.115, 0.800	0.327	-0.193, 1.042
Portugal	0.206	-0.304, 0.762	0.180	-0.373, 0.954
Spain	0.292	-0.226, 0.800	0.323	-0.273, 1.039
Sweden	0.234	-0.230, 0.716	0.171	-0.294, 0.678
Switzerland	-0.477	-0.885, -0.030	-0.956	-2.001, -0.196
UK	-0.027	-0.462, 0.413	0.038	-0.322, 0.565

b) x = 6

c) x =

Figure 1. Average annual nominal exchange rate depreciation and relative price inflation from 1870 to 2013

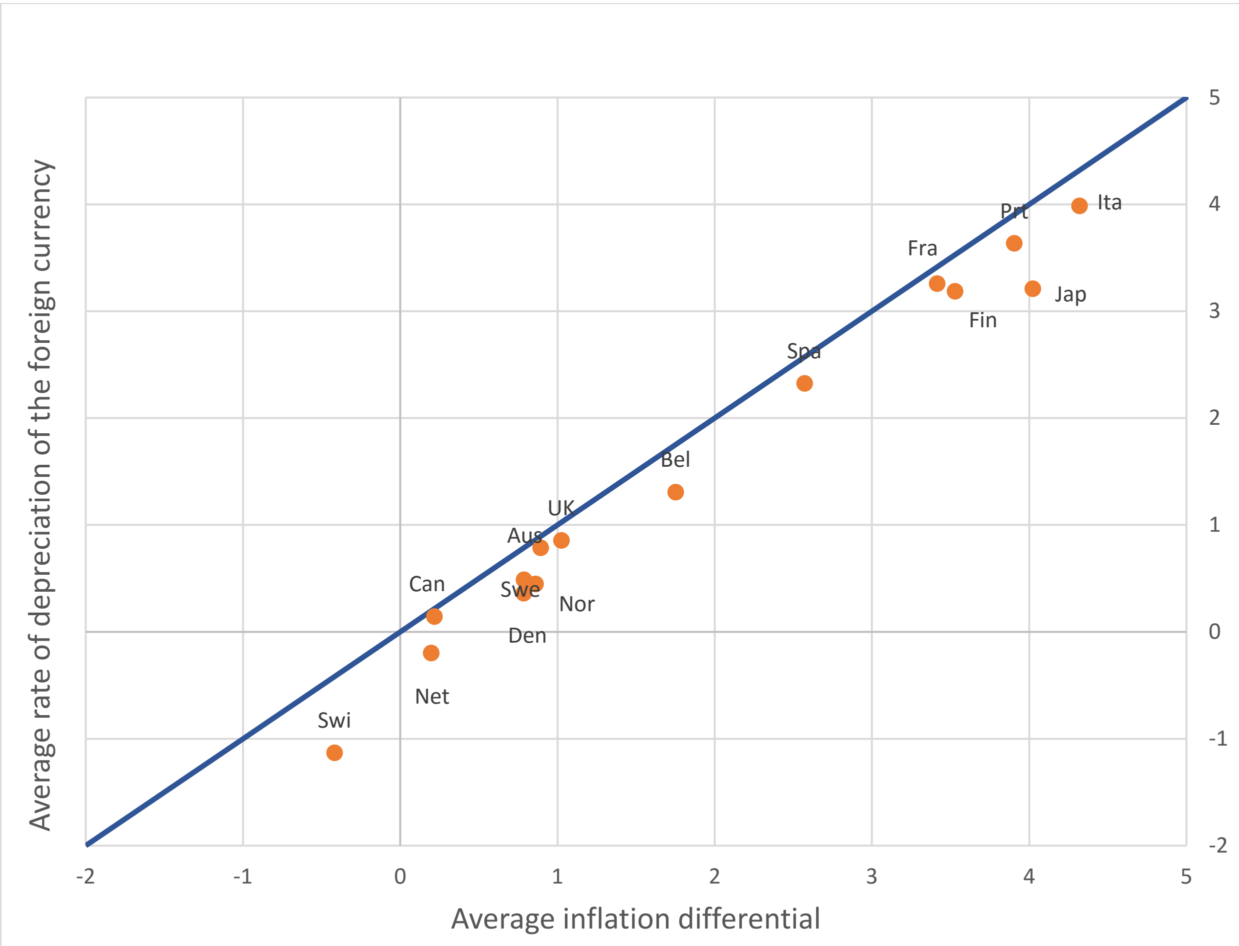
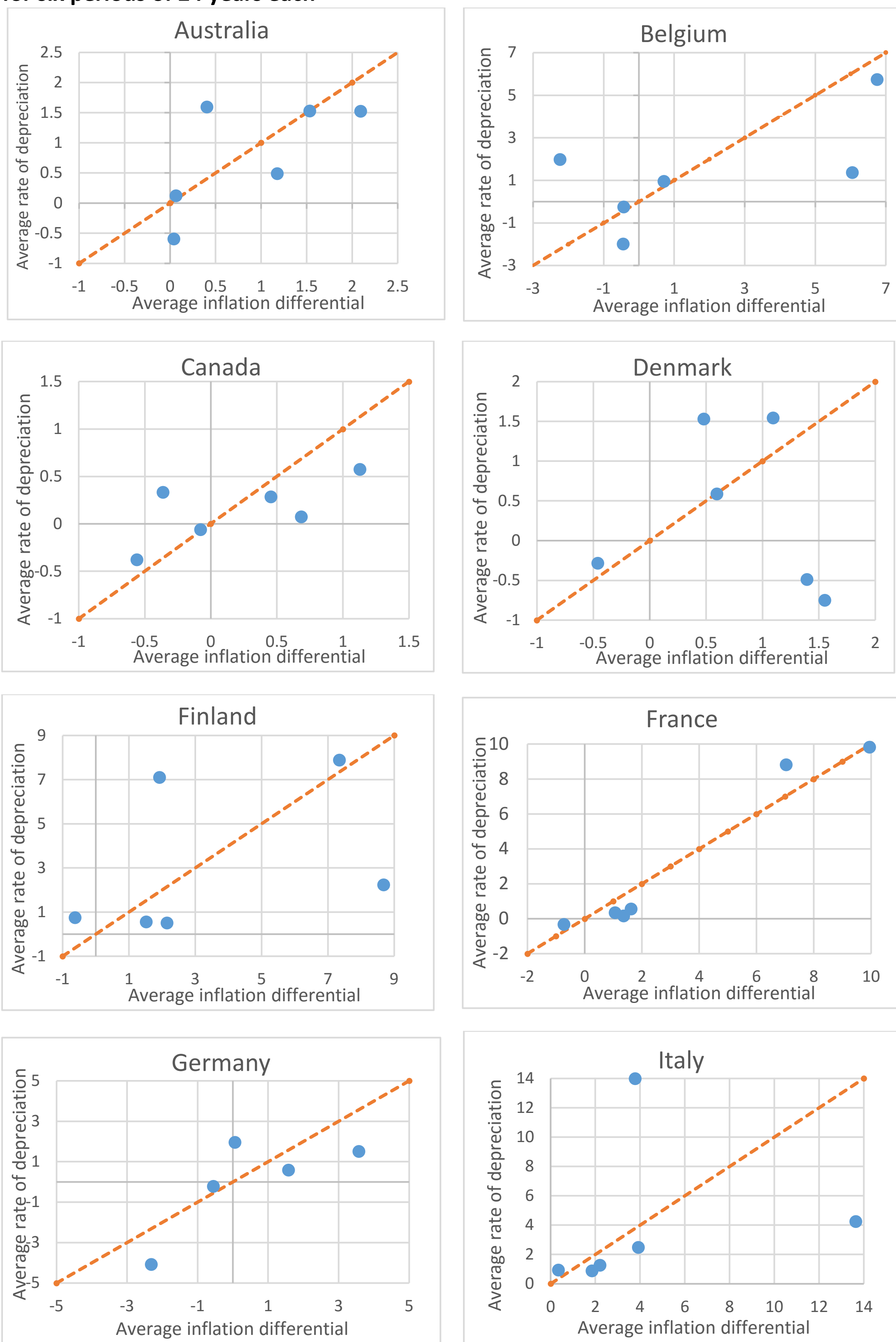
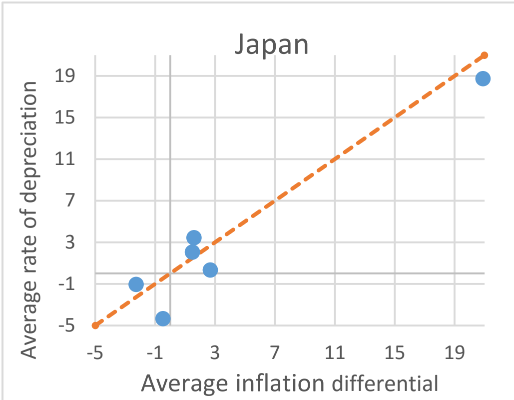
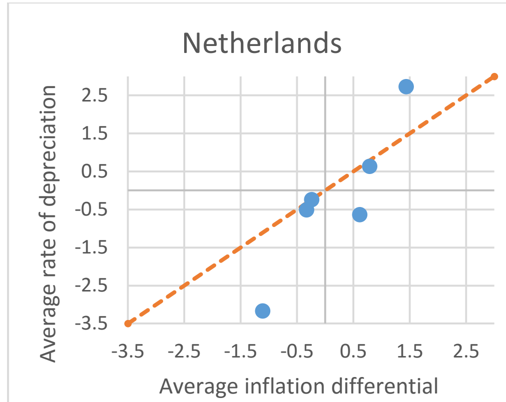
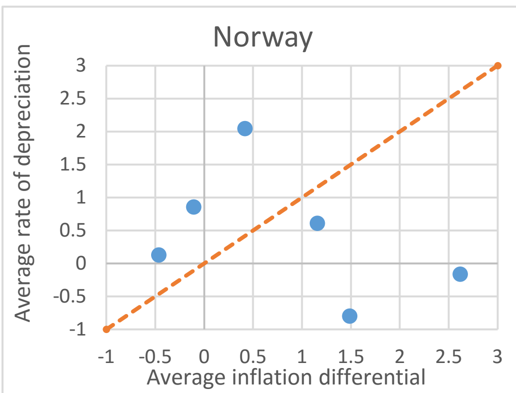


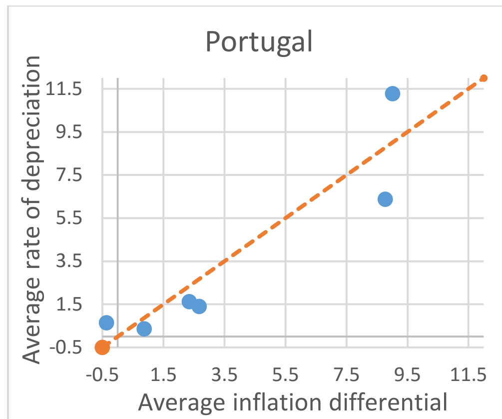
Figure 2. Average annual nominal exchange rate depreciation and relative price inflation for six periods of 24 years each

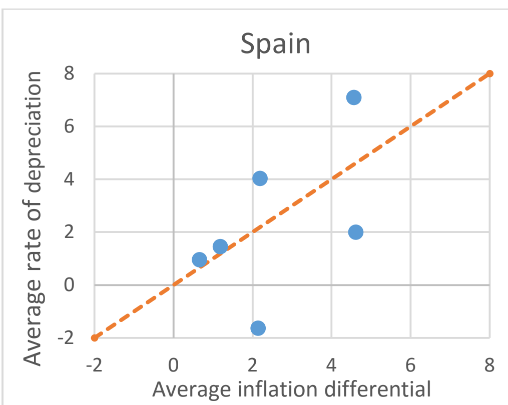


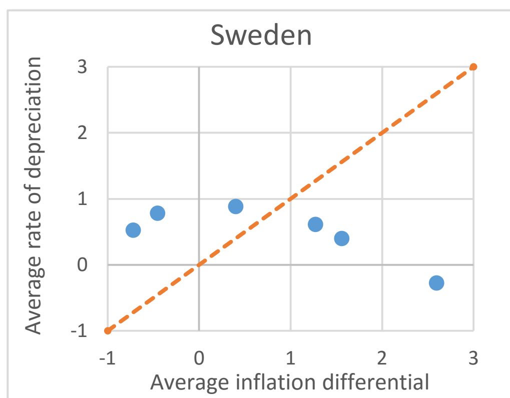

Japan
Average rate of depreciation
19
15
11
7
3
-1
-5
-5
-1
3
7
11
15
19
Average inflation differential

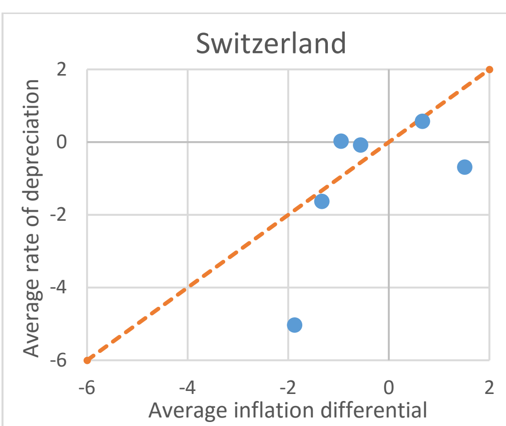

Netherlands
Average rate of depreciation
2.5
1.5
0.5
-0.5
-1.5
-2.5
-3.5
-3.5
-2.5
-1.5
-0.5
0.5
1.5
2.5
Average inflation differential

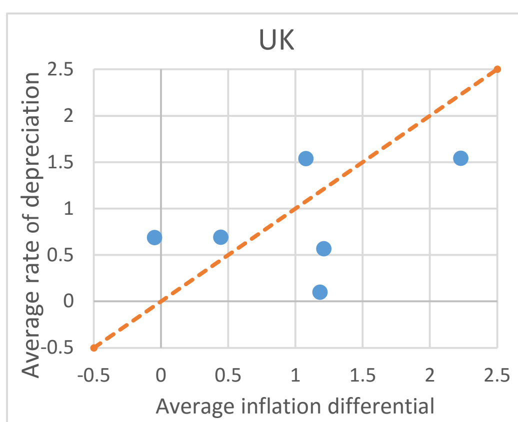

Norway
Average rate of depreciation
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1
-0.5
0
0.5
1
1.5
2
2.5
3
Average inflation differential


Portugal
Average rate of depreciation
11.5
9.5
7.5
5.5
3.5
1.5
-0.5
-0.5
1.5
3.5
5.5
7.5
9.5
11.5
Average inflation differential


Spain
Average rate of depreciation
8
6
4
2
0
-2
-2
0
2
4
6
8
Average inflation differential


Sweden
Average rate of depreciation
3
2
1
0
-1
-1
0
1
2
3
Average inflation differential


Switzerland
Average rate of depreciation
2
0
-2
-4
-6
-6
-4
-2
0
2
Average inflation differential


UK
Average rate of depreciation
2.5
2
1.5
1
0.5
0
-0.5
-0.5
0
0.5
1
1.5
2
2.5
Average inflation differential