Long Run Purchasing Power Parity: Cassel or Balassa-Samuelson?

David H. Papell and Ruxandra Prodan

University of Houston

November 2003

We use long-horizon real exchange rate data for 16 industrialized countries to investigate two alternative versions of Purchasing Power Parity (PPP): reversion to a constant mean in the spirit of Cassel and reversion to a constant trend in the spirit of Balassa and Samuelson. We develop unit root tests that both account for structural change and maintain a long run mean or trend. Using conventional tests, we find evidence of some variant of PPP for 9 of the 16 countries. With the restricted tests, we find evidence for 5 additional countries. The Cassel version of PPP is supported for 10 countries and the Balassa-Samuelson version is supported for 4 countries. The simulation results show good size and power of the restricted tests.

We are grateful to participants in seminars at Emory University, the University of Michigan, Texas Camp Econometrics VIII and the 2002 Southern Economic Association Meetings for helpful comments and discussions. Papell thanks the National Science Foundation for financial support.

Correspondence to:
Department of Economics, University of Houston, Houston, TX 77204-5019.
Ruxandra Prodan, tel. (713) 743-3817, fax (713) 743-3798, email: rprodan@mail.uh.edu
David Papell, tel. (713) 743-3807, fax (713) 743-3798, email: dpapell@mail.uh.edu

1. Introduction

Purchasing Power Parity (PPP) is one of the oldest and most studied topics in international economics. As articulated by Cassel (1918), the absolute version of PPP postulates that the relative prices (in different currencies and locations) of a common basket of goods will be equalized when quoted in the same currency. The relative version of PPP, emphasizing arbitrage across time rather than across space, is that the exchange rate will adjust to offset inflation differentials between countries.[1] While Cassel understood the possibility that the exchange rate might transitorily diverge from PPP, he viewed the deviations as minor.[2] Modern versions of PPP, recognizing the importance of slow speeds of adjustment, define PPP as reversion of the real exchange rate to a constant mean.

Drawing on ideas from Ricardo and Harrod, Balassa (1964) and Samuelson (1964) drew attention to the fact that divergent international productivity levels could, via their effect on wages and home good prices, lead to permanent deviations from Cassel's absolute version of PPP. They linked PPP, exchange rates and intercountry real-income comparisons, arguing that the absolute version of PPP is flawed as a theory of exchange rates. Assuming that PPP holds for traded goods, their argument is based on the fact that productivity differentials between countries determine the domestic relative prices of non-tradables, leading in the long run to trend deviations from PPP.[3] Obstfeld (1993) uses these ideas to develop a model in which real exchange rates contain a pronounced deterministic trend.

The tension between these two approaches is evident in recent studies of PPP. Modern research on PPP, taking into account the possibility of slow speeds of reversion, finds evidence of PPP if the unit root null can be rejected in favor of a stationary alternative for real exchange rates. The question remains, however, whether the stationary alternative should be level stationarity, reversion to a constant mean, or trend stationarity, reversion to a constant trend. These issues are of central importance in studying long-horizon real exchange rates. In the

[1] Since data on price indexes for different countries do not measure a common basket of goods, empirical work on PPP usually tests relative, rather than absolute, PPP.
[2] He identified three groups of disturbances: actual and expected inflation or deflation, new hindrances to trade and shifts in international movements of capital. But even if the disturbances are recognized, their quantitative effect on deviations from PPP is seen as "confined within rather narrow limits".
[3] There are several studies which point out the limitations of this theory: Asea and Mendoza (1994) found little evidence to support the proposition that deviations from PPP reflect differences in the relative prices of non-tradables, Canzoneri (1999) at al. found less favorable evidence on purchasing power parity in traded goods, and Fitzgerald (2003) argues that the classic relationship between productivity and relative price levels is modified by adding the terms-of-trade effects to the model.

context of studying real exchange rates during the post-1973 floating exchange rate regime, productivity differentials and the resultant possibility of trending assumes lesser importance.

Purchasing Power Parity with long-horizon real exchange rates has been investigated using both approaches. In the Cassel tradition, Abuaf and Jorion (1990) and Lothian and Taylor (1996) find evidence of long-run PPP by rejecting unit roots in favor of level stationary real exchange rates using Augmented-Dickey-Fuller (ADF) tests. Nonetheless, the convergence to PPP is very slow, with half-lives of PPP deviations between 3 and 5 years.[4] Murray and Papell (2002) and Lopez, Murray, and Papell (2002), using median unbiased techniques to measure persistence and calculating confidence intervals as well as point estimates, find even longer half-life estimates.

In the Balassa-Samuelson tradition, Taylor (2002) creates a long-horizon real exchange rate data set for 17 industrialized and 4 Latin American countries. He finds that long-run PPP can be supported in almost all cases using the Elliott, Rothenberg and Stock (1996) generalized-least-squares version of the Dickey-Fuller (DF-GLS) test with allowance for a deterministic trend. Cuddington and Liang (2000) and Lothian and Taylor (2000) investigate the implications of allowing for trending in the Lothian and Taylor (1996) data.

Engel (2000) decomposes the real exchange rate into two processes, one stationary (relative price of traded goods) and the other non-stationary (relative price of non-traded goods). He shows that, due to size bias, the non-stationary component of the real exchange rate may not be detected in tests for long run PPP. Ng and Perron (2002), following the same argument as Engel, show that even if in some cases testing for unit root is difficult, improvements can be made by minimizing Type I error and maximizing the power of the tests.[5]

Starting with Perron (1989), much research has been conducted on testing for unit roots in the presence of a one-time change in the mean, trend, or both the mean and the trend of economic time series. This is important in tests for PPP because, if there is a one-time change in the mean, long-run PPP does not hold. The situation becomes more complicated when, as in Lumsdaine and Papell (1997), multiple structural changes are allowed. Suppose that the real exchange rate is subject to two changes in the mean. If the changes are offsetting, the series returns to a constant mean and long-run PPP holds. If the changes are not offsetting, either because they act in the same direction or because they act in opposite directions but are of

[4] The combination of slow speed of convergence to PPP and high short-term volatility of real exchange rates is called the "The Purchasing Power Parity Puzzle" by Rogoff (1996).
[5] In order to minimize size distortions they propose a new information criteria, MAIC, to select the autoregressive lag length

different magnitude, the series does not return to a constant mean and long-run PPP does not hold.

Dornbusch and Vogelsang (1991) argue that a “qualified” version of purchasing power parity can still be claimed in the presence of a one-time shift in the mean level of the real exchange rate that is determined exogenously. They interpret their findings as supporting the Balassa-Samuelson model. Hegwood and Papell (1998) formalize and generalize this idea, allowing for multiple structural changes that are determined endogenously. They argue that real exchange rates are level stationary, but around a mean which is subject to structural change, and show that reversion to a changing mean is much faster than reversion to a fixed mean.[6]

The first step in our investigation is to use tests that do not incorporate structural change. In order to avoid confusion between differing concepts of PPP, we call “Purchasing Power Parity” (PPP) the rejection of the unit root null hypothesis in favor of an alternative hypothesis of level stationarity in a model that does not incorporate a time trend. The data is from Taylor (2002), and consists of 107 to 129 years of real exchange rates for 16 industrialized countries with the United States dollar as the numeraire currency.

We call ”Trend Purchasing Power Parity” (TPPP) the rejection of the unit root null in favor of a trend stationary alternative in a model that incorporates a time trend. Using conventional Augmented Dickey-Fuller (ADF) unit root tests, we find evidence of PPP by rejecting the unit root null (at the 5% level) for only 8 out of 16 countries.[7] Incorporating time trends, we also find evidence of TPPP for 7 countries. Combining the two tests, we find evidence of either PPP or TPPP for 9 out of 16 countries. This is weaker evidence than found by Taylor (2002). The reason for our weaker evidence is that we use lag selection methods that provide better size and power properties for the tests, not that our ADF tests have lower power than Taylor’s DF-GLS tests.

Tests of the unit root hypothesis in the presence of a one-time change in the intercept have been developed for both non-trending data, as in Perron and Vogelsang (1992), and trending data, as in Perron (1997) and Vogelsang and Perron (1998). Subsequently we test the unit root hypothesis in real exchange rates allowing for a possible shift in the intercept of the trend function, considered to occur at an unknown time. As an extension, we propose tests that extend the previous models to incorporate two endogenous break points.

[6] Hegwood and Papell (1998) use exchange rates for five countries with the US dollar for 1900 to 1990, plus the two exchange rates used by Lothian and Taylor (1996).

[7] We use the term “country” as shorthand for “real exchange rate with the United States”.

Following our earlier terminology, we call "Qualified Purchasing Power Parity" (QPPP) the rejection of the unit root hypothesis in favor of an alternative hypothesis of regime-wise level stationarity (level stationarity after allowing for one or two changes in the intercept). We call "Trend Qualified Purchasing Power Parity" (TQPPP) the rejection of the unit root hypothesis in favor of an alternative hypothesis of regime-wise trend stationarity (trend stationarity after allowing for one or two changes in the intercept).

We first report the results of tests that do not impose restrictions on the breaks. Allowing for one or two structural changes, we find evidence of QPPP for 13 out of 16 countries and of TQPP for 11 out of 16 countries (at the 5% level). In no case, however, do we find evidence of TQPPP for a country for which we do not find evidence of QPPP and so, combining the two tests, we find evidence of either QPPP or TQPPP for 13 countries.

These results are not necessarily a step forward towards PPP. They do not impose either the Cassel alternative of a constant mean or the Balassa-Samuelson alternative of a constant trend. In order to test for PPP while allowing for structural change, we develop unit root tests that restrict the coefficients on the dummy variables that depict the breaks to produce a constant mean or trend in the long run. We therefore account for structural change but still maintain the long-run PPP or TPPP hypothesis. We call the model for non-trending data "PPP restricted structural change" and the model for trending data "Trend PPP restricted structural change". It is important to understand that, in contrast with the tests that allow for QPPP and TQPPP alternatives, rejection of the unit root null in favor of the restricted structural change alternatives provides evidence of PPP or TPPP[8]

We reject the unit root hypothesis in favor of PPP restricted structural change for 7 countries (at the 5% level). When the alternative is trend PPP restricted structural change, we reject the unit root null for 12 countries. Combining these tests with the ADF tests for PPP and TPPP, Canada and the Netherlands are the only countries for which there is no evidence of any variant of PPP[9]. In case of Canada, where the unit root null cannot be rejected in favor of any of our alternative hypotheses, one possibility for the lack of evidence of PPP is because of the large depreciation of the Canadian dollar against the US dollar at the end of the sample. For the Netherlands, where the unit root null can only be rejected in favor of QPPP with one break, there

[8] A related test was developed by Papell (2002) to account for the large appreciation and depreciation of the dollar in the 1980s. Using panel methods for post-1973 data and imposing a "PPP restricted broken trend" constraint he provides strong evidence of PPP.

[9] For Germany and Australia, we found evidence of PPP and TPPP, respectively, but did not find evidence of PPP or TPPP restricted structural change.

is a real appreciation of the florin against the US dollar in the early 1970s that has not been reversed.

We have presented evidence that, using tests that do not allow for structural change, PPP or TPPP can be found for only 9 out of 16 countries. With tests that both allow for structural change and impose parity restrictions, evidence of PPP or TPPP can be found for 14 out of 16 countries. In order to evaluate these results, we conduct simulations to evaluate the power of the no break, two breaks, and restricted two breaks tests under various alternative hypotheses.

The simulations produce several interesting results. First, the power of the ADF tests without structural change is miniscule when the data is generated by a process that incorporates structural change, even if the process is consistent with PPP or TPPP (the breaks are of equal and opposite sign).[10] This implies that the rejections using ADF tests do provide evidence of PPP or TPPP without structural change. Second, the tests with restricted structural change have very high power when the process has breaks of equal and opposite sign, moderate power when the process has no breaks, and very low power when the process has breaks that are not consistent with PPP or TPPP. This implies that the additional rejections from the tests with restricted structural change constitute evidence of PPP and TPPP beyond what can be obtained with the ADF tests. As a further check on our findings of PPP or TPPP restricted structural change, we compare the break dates and coefficients from the restricted to those from the unrestricted structural change models. Put together, the results in this paper provide strong evidence of PPP or TPPP for 14 out of 16 long-run real exchange rates.

2. Testing for PPP and TPPP with and without Structural Change

The purpose of the paper is to analyze long-run purchasing power parity among industrialized countries. We use annual nominal exchange rates and price indices. The latter are measured as consumer price deflators or GDP deflators, depending on their availability. The data was obtained from Taylor (2002) and was updated by increasing the sample (using International Financial Statistics data from 2001). The data covers a set of 16 industrialized countries and US as the base country, starts from 1870 for the longest series, and ends in 1998.[11]

[10] This result varies with the break size and the level of persistence of the series; the power of ADF test becomes higher if the size of the break is very small and/or the level of persistence is very low.

[11] The countries are Australia, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom and United States. The number of years varies across countries, but we have a complete cross section of 17 countries between 1892 and 1996.

Under purchasing power parity (PPP), the real exchange rate displays long-run mean reversion. The real dollar exchange rate is calculated as follows:

$$q = e + p^* - p \tag{1}$$

where q is the logarithm of the real exchange rate, e is the logarithm of the nominal exchange rate (the dollar price of the foreign currency) and p_t and $p_t{}^*$ are the logarithms of the US and the foreign price levels, respectively.

We proceed in three steps: First, we test for PPP and TPPP without allowing for structural change. Second, we test for unit roots while allowing for structural change, but do not impose PPP or TPPP. Third, we test for unit roots in the presence of PPP or TPPP restricted structural change.

2.1. Conventional tests for stationarity

We first check for the stationarity of the data using the Augmented Dickey-Fuller (ADF) unit root test, which regresses the first difference of the logarithm of the real exchange rate on a constant, its lagged level and k lagged first differences:

$$\Delta q_t = \mu + \alpha q_{t-1} + \sum_{i=1}^{k} c_i \Delta q_{t-i} + \varepsilon_t \tag{2}$$

The number of lags is chosen by using the general-to specific recursive t-statistic procedure suggested by Campbell and Perron (1991) and Ng and Perron (1995). We set the maximum value of k equal to 8, and use a critical value of 1.645 from the asymptotic normal distribution to assess significance.[12]

The null hypothesis of a unit root is rejected in favor of the alternative of level stationarity if α is significantly different from 0. In this case, the rejection of the unit root null provides evidence of Purchasing Power Parity (PPP). If a time trend is included in Equation (2), the null hypothesis of a unit root is rejected in favor of the alternative of trend stationarity if α is

[12] The choice of the maximum value of k is somewhat arbitrary. On the one hand, one would like a large value to have as unrestricted a procedure as possible. On the other hand, a large value of k max yields problems of multicolliniarity in the data and also substantial loss of power.

significantly different from 0. In this case, rejection of the unit root null provides evidence of Trend Purchasing Power Parity (TPPP).

The results of the ADF tests are reported in Table 1. We find evidence of PPP by rejecting the unit root null for 8 out of 16 countries at the 5% level, Belgium, Germany, Finland, France, Italy, Norway, Spain, and Sweden, and for 3 additional countries at the 10% level. Incorporating time trends, we find evidence of TPPP by rejecting the unit root null for 7 countries at the 5% level, Australia, Belgium, Finland, France, Italy, Norway, Sweden and for 3 additional countries at the 10% level. Combining the two tests, we find evidence of either PPP or TPPP by rejecting the unit root null for 9 countries at the 5% level and 2 additional countries at the 10% level. Since ADF tests that do not incorporate a time trend have very low power to reject unit roots with trending data, we interpret these results as providing evidence of PPP for the 8 countries for which the unit root null is rejected in favor of the PPP alternative, and of TPPP for one country, Australia, for which the unit root null is rejected in favor of the TPPP, but not the PPP, alternative.[13]

Taylor (2002) reports stronger evidence of PPP. Using the Elliott, Rothenberg and Stock (1996) DF-GLS test, he rejects the unit root null for 10 out of 16 countries at the 5% level and for 3 additional countries at the 10% level with the U.S. dollar as the numeraire currency. He also finds stronger evidence of TPPP, rejecting the unit root null for 8 out of 16 countries at the 5% level and for 3 additional countries at the 10% level. Combining the two tests, he finds evidence of either PPP or TPPP for 11 countries at the 5% level and 4 additional countries at the 10% level. He also reports marginally stronger rejections with ADF tests than we find, rejecting the unit root null in favor of either PPP or TPPP for 9 out of 16 countries at the 5% level and for 3 additional countries at the 10% level.[14]

The reason for the differences in the results comes from differences in lag selection procedures. Taylor uses a Lagrange Multiplier technique, which tends to produce short lag lengths. As shown by Ng and Perron (1995, 2001), techniques that produce short lag lengths have low power for ADF tests and are badly sized for DF-GLS tests. Using the techniques recommended by Ng and Perron to produce tests with good size and power, general-to-specific

[13] West (1987) shows that ADF tests that do not include a time trend have zero asymptotic power if the series is trend stationary. Taylor (2002) indicates that at least for some series a deterministic term may be present, the trend being sizeable: 1.5% per annum in the case of Japan, 0.74% per annum for Switzerland, but small in all other cases.
[14] The comparisons with Taylor's results are for industrialized countries with the U.S. dollar as numeraire. He also reports results with four Latin American countries and, for all countries, with a "world basket" as numeraire. While we add two years to Taylor's data, none of the comparisons are affected by using his original sample.

lag selection for ADF tests and MAIC lag selection for DF-GLS tests, we reject approximately as much as Taylor using ADF tests and less often using DF-GLS tests.

Lopez, Murray and Papell (2002), using DF-GLS tests with MAIC lag selection, reject the unit root null in favor of the PPP alternative for 9 out of 16 countries at the 5% level. We conducted DF-GLS tests with MAIC lag selection and TPPP as the alternative hypothesis, and could only reject the unit root null for 5 out of 16 countries at the 5% level and for 4 additional countries at the 10% level. Combining the two tests, evidence of either PPP or TPPP using DF-GLS tests with MAIC lag selection can be found for 9 countries at the 5% level and 2 additional countries at the 10% level. With up to 129 years of data, it should not be surprising that the power gains from using DF-GLS tests rather than ADF tests do not make much difference in practice. Whether we use ADF tests with general-to-specific lag selection or DF-GLS tests with MAIC lag selection, evidence of either PPP or TPPP can be found for only 9 of the 16 countries at the 5% level.

2.2 Tests for a unit root in the presence of structural change

As Campbell and Perron (1991) emphasized, nonrejection of the unit root hypothesis may be due to the misspecification of the deterministic components included as regressors. Unit root tests that ignore structural change could fail to provide evidence of PPP when it actually holds outside of the structural shift. We investigate the unit root hypothesis in real exchange rates, but not PPP or TPPP, by using previously developed tests for a unit root in the presence of one break and by developing tests for a unit root with two breaks. The intuition that motivates the tests is to treat the breaks as being determined outside the data generating process.[15]

Tests for a unit root in a non-trending time series characterized by a single structural change in its level are developed by Perron and Vogelsang (1992). The possible changes are considered to occur at an unknown time. We consider an Additive Outlier type (AO) model to model changes that occur instantaneously. Analyzing long-run real exchange rates, we mainly draw the real exchange rates from a nominal fixed exchange rate regime, where devaluations and reevaluations, especially following failed attempts to defend currencies, can lead to discrete jumps (intercept changes).

[15] While there is no theoretical reason to restrict attention to one or two breaks, practical considerations involving computing time for simulations and calculating critical values precluded considering additional breaks.

The AO model is estimated using a two-step process. For a value of the break point TB, with .10T<TB<.90T (where T is the sample size), the deterministic part of the series is removed using the following regression:

$$q_t = \mu + \gamma DU + \tilde{z}_t \tag{3}$$

where DU = 1 if T>TB, 0 otherwise. The 10% trimming is used to avoid finding spurious "breaks" at the beginning and end of the sample. The unit root test is then performed using the t statistic for α = 0 in the regression:

$$\Delta\tilde{z}_t = \sum_{i=0}^{k} \omega_i D(Tb)_{t-i} + \alpha\tilde{z}_{t-1} + \sum_{i=1}^{k} c_i \Delta\tilde{z}_{t-i} + \varepsilon_t \tag{4}$$

where D(TB)t = 1 if t = TB+1 and 0 otherwise. The inclusion of k+1 dummy variables is needed to ensure that t statistic on α is invariant to the value of truncation lag parameter k. The recursive procedure of selecting the truncation lag parameter k starts with kmax =8 and it is repeated until the last lag is significant (use a critical value of 1.645).

The break date, TB is chosen to minimize the t-statistic on α. Statistics are computed for all break dates, taking into account the trimming. The chosen break is that for which the maximum evidence against the unit root null, in the form of the most negative t-statistic on α, is obtained.

Tests for a unit root in a trending time series characterized by a single structural change are developed by Perron (1997) and Vogelsang and Perron (1998). Including a time trend, we follow the procedure described above and perform unit root tests that allow shifts in the intercept at an unknown time.[16]

As previously, the AO model is estimated using a two-step process. The deterministic part of the series is removed using the following regression:

$$q_t = \mu + \beta t + \gamma\, DU + \tilde{z}_t, \tag{5}$$

[16] There are three possible models in case of trending data: intercept shift, intercept and slope shift and a slope shift. We did calculations for all but we didn't find more rejections or other countries than in the model that allows only for changes in the intercept.

where 10% trimming is used to avoid finding spurious breaks. The unit root test is then performed using the t statistic for $\alpha = 0$ in the regression described by Equation (4). The unit root hypothesis is tested as in the previous case.

Critical values were computed using Monte Carlo methods. We generate a unit root series (without structural change) with 129 observations (the maximum size of the sample), using an AR (1) model with *iidN*(0,1) innovations. The AO model is estimated as described above, with the test statistic being the t-statistic on α in Equation (4). The critical values for the finite sample distributions are taken from the sorted vector of 5000 replicated statistics.[17]

The results of the tests for a unit root in the presence of one structural change are reported in Table 2. We find evidence of Qualified Purchasing Power Parity (QPPP) by rejecting the unit root null for 11 out of 16 countries at the 5% level, Australia, Belgium, Denmark, Finland, France, Italy, Netherlands, Norway, Spain, Sweden and Switzerland and for one additional country at the 10% level. Incorporating time trends, we find evidence of Trend Qualified Purchasing Power Parity (TQPPP) by rejecting the unit root null, for 7 countries at the 5% level, Belgium, Finland, France, Italy, Norway, Sweden and Switzerland, and for 5 additional countries at the 10% level. Combining the two tests, we find evidence of either QPPP or TQPPP by rejecting the unit root null for 11 countries at the 5% level.

We proceed to extend the AO model of Perron and Vogelsang (1992), for non-trending data, to incorporate two endogenous break points.[18] Following the previous testing procedures, the AO model is estimated using a two-step process. For values of the break points TB1 and TB2 with .10T<TBi<.90T (where T is the sample size and i = 1,2), the deterministic part of the series is removed using the following regression:

$$q_t = \mu + \gamma_1 DU1_t + \gamma_2 DU2_t + \tilde{z}_t \tag{6}$$

where $DU1_t = 1$ if T>TB1, 0 otherwise and $DU2_t = 1$ if T>TB2, 0 otherwise. The unit root test is then performed using the t statistic for $\alpha = 0$ in the regression:

$$\Delta\tilde{z}_t = \sum_{i=0}^{k} \omega_i D(Tb1)_{t-i} + \sum_{i=0}^{k} \omega_i D(Tb2)_{t-i} + \alpha\tilde{z}_{t-1} + \sum_{i=1}^{k} c_i \Delta\tilde{z}_{t-i} + \varepsilon_t \tag{7}$$

[17] We experimented by computing data specific critical values for several countries, and the results were unaffected.
[18] Lumsdaine and Papell (1997) develop a unit root test that allows for two endogenously determined break points in the context of an innovational outlier (IO) model for trending data, where the structural change is assumed to occur gradually. This assumption is more appropriate for macroeconomic aggregates than for real exchange rates.

where D(TBi)t = 1 if t = TBi+1 and 0 otherwise. Statistics are computed for all possible combinations of break dates, taking in account the trimming and not allowing breaks to occur in consecutive years.

Finally, we extend the AO model of Vogelsang and Perron (1998), for trending data, to allow for two breaks. We follow the procedure described above and perform unit root tests that allow shifts in the intercept at an unknown time. The deterministic part of the series is removed using the following regression:

$$q_t = \mu + \beta t + \gamma_1 DU1_t + \gamma_2 DU2_t + \tilde{z}_t \tag{8}$$

The unit root test is then performed using the t statistic for α = 0 in the regression described by Equation (7). The unit root hypothesis is tested as in the previous case.

The results of the tests for a unit root in the presence of two structural changes are reported in Table 3. We find evidence of Qualified Purchasing Power Parity (QPPP) by rejecting the unit root null for 12 out of 16 countries at the 5% level, Australia, Belgium, Denmark, Finland, France, Italy, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom, and for one additional country at the 10% level. Incorporating time trends, we find evidence of Trend Qualified Purchasing Power Parity (TQPPP) by rejecting the unit root null, for 11 countries at the 5% level, Belgium, Denmark, Finland, France, Italy, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom, and for one additional country at the 10% level. Using both tests, we find evidence of either QPPP or TQPPP by rejecting the unit root null for 12 countries at the 5% level. We conclude by combining the results of the tests with one and two breaks. There is one country, the Netherlands, for which evidence of QPPP is found with one, but not two, breaks. We therefore reject the unit root null in favor of either the QPPP or the TQPPP alternative for 13 countries at the 5% level.

2.3 Tests for PPP in the presence of restricted structural change

We first presented evidence of PPP (or TPPP) for 9 out of 16 countries using a conventional ADF test. Then, using an AO model and allowing for one or two intercept changes we found evidence of QPPP (or TQPP) for 13 countries. Neither of these tests, however, investigates the possibility that a series may experience both structural change and reversion to the mean (or trend). In order to test for PPP (or TPPP) while allowing for structural change, we develop unit root tests that restrict the coefficients on the dummy variables that depict the breaks to produce a long-run constant mean or trend. We estimate AO models that maintain the long run PPP hypothesis by removing the deterministic part of the series using the following regression:

$$q_t = \mu + \gamma_1 DU1_t + \gamma_2 DU2_t + \tilde{z}_t \tag{9}$$

subject to the restriction: $\gamma_1 + \gamma_2 = 0$ (10)

where 10% trimming is used to avoid finding spurious breaks. The restriction in Equation (10) is that the coefficients on the breaks are of equal and opposite sign. This imposes the PPP hypothesis because the mean following the second break is restricted to equal the mean prior to the first break. The unit root test is then performed using the t statistic for $\alpha = 0$ in the regression described by Equation (7). The unit root hypothesis is tested as in the previous case.

Next, we estimate the AO model which includes a time trend, described by Equation (8), subject to the same restrictions (10). This imposes the TPPP hypothesis because the trend following the second break is restricted to equal the trend prior to the first break. We follow the same procedure as before to choose the breaks and test the unit root hypothesis. Critical values for the restricted models are calculated using the same method as in the case of the AO model with two structural breaks. Because of the restrictions, their values are lower than in the case of two unrestricted breaks.

The results of the tests with restricted structural change are reported in Table 4. We find evidence of PPP Restricted Structural Change by rejecting the unit root null for 7 out of 16 countries at the 5% level, Finland, France, Italy, Norway, Portugal, Spain and the United Kingdom, and for two additional countries at the 10% level. Incorporating time trends, we find evidence of Trend PPP Restricted Structural Change by rejecting the unit root null for 12 countries at the 5% level, Belgium, Denmark, Finland, France, Italy, Japan, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom, and for one additional country at the 10% level. Combining the two tests, we find evidence of either PPP or TPPP Restricted Structural Change by rejecting the unit root null for 12 countries at the 5% level.

Australia, Canada, Germany and the Netherlands are the countries for which we do not find evidence of either PPP or TPPP Restricted Structural Change. Among these countries, we previously found evidence of PPP for Germany and of TPPP for Australia by using ADF tests.

Combining the tests, we find evidence of some variant of PPP for 14 of the 16 countries. In addition, we find evidence of QPPP for the Netherlands. Canada is the only country for which there is no evidence of some variant of stationarity.

3. Size and power of univariate unit root tests with structural change

The central empirical result of the paper is that unit root null can be rejected in favor of a variant of PPP for 14 out of 16 countries, 5 more rejections than from a conventional ADF test. In order to determine whether these rejections constitute evidence of PPP, TPPP, PPP restricted structural change, or trend PPP restricted structural change, we study the finite sample performance of the newly developed tests and proceed to investigate their size and power.

3.1 Construction of the size and power simulations

The simulation experiments address the following issues: (a) comparison of various tests properties (ADF, AO model allowing for two structural changes and AO restricted model), (b) power of AO models as a function of the level of persistence and the magnitude of the break, (c) power of the PPP restricted tests on trending data and trend PPP restricted tests on non-trending data.[19]

Within the Monte Carlo experiments we consider the following two data generating processes:

$$q_t = \mu + (\beta t) + \alpha q_{t-1} + \varepsilon_t \tag{11}$$

$$q_t = \mu + (\beta t) + \alpha q_{t-1} + \gamma_1 DU1_t + \gamma_2 DU2_t + \varepsilon_t \tag{12}$$

The power of the unit root tests can be investigated by constructing experiments with artificial data under a true alternative hypothesis where the real exchange rate is stationary (or trend stationary) without structural change (11) and, also, allowing for two changes in the intercept (12). We perform unit root tests on these constructed series, tabulating how often the unit root null is (correctly) rejected.

The two different sets of data are generated based on different assumptions: we specify α = 0.7, 0.8 and 0.9 for first generated sample (11) and α = 0.5, 0.6, 0.7 and 0.8 for the second

[19] We do not report simulation results for AO models with one break because they are similar to those with two breaks in the same direction.

generated sample (12). This reflects the range of values of α reported in Tables 1 – 4 (with and without structural changes).

In the case of generated data including two structural changes we consider cases where the breaks are equal and have either the same sign or the opposite sign. We account for breaks with a magnitude of 0.1, 0.3 and 0.5, which covers most of the cases found in our data. The timing of the breaks is set at the 1/3 and 2/3 of the sample. In all of the cases the sample size is T = 129 with 10% trimming and 1000 replications are used with $e_t = iidN(0,1)$. We also use the finite sample critical values previously calculated and report results for tests of nominal size of 10%, 5% and 1%.

3.2 Simulation results

To address these issues it is useful to start with stationary data generating processes that do not contain structural change under the alternative hypothesis. In this section, we report power results for tests with a nominal size of 5% and α = 0.7, 0.8 and 0.9 (detailed results for different nominal sizes and values of α are presented in Table 5). First, as the errors become more persistent, the power decreases: The ADF test has good power when the data generating process is stationary with α = 0.7 and 0.8 but lower power when α = 0.9. This is in accord with previous research, considering the span and persistence of the data.[20] As expected, tests for QPPP that incorporate two structural changes have less power than the ADF test. Tests for PPP restricted structural change have lower power than the ADF tests (except in the case of low persistence) but higher power than QPPP test.

Generating trend stationary data processes and applying the previous tests, including a time trend, we obtained fairly similar results with the previous ones. However, the power of the tests, in most of the cases, is lower than in the above case. As in the case of non-trending data, the power of trend-restricted tests exceeds the power of ADF tests with low persistence.

Consider now experiment (2) in which we generate stationary data and allow for two changes in the intercept (Table 6). We start by looking at cases where the breaks occur in the same direction (inconsistent with PPP). The test for QPPP with two structural changes has very good power, except for the case with high persistence (α = 0.8). For low levels of persistence

[20] A coefficient of α of .8 corresponds to a half-life with annual data of 3.11 years.

(α = 0.5, 0.6) there is generally a small decrease in power as the breaks become larger.[21] For higher levels of persistence (α = 0.7, 0.8) the simulation results show evidence of non-monotonic power. The power first decreases and then increases as the size of the break rises.[22]

Because the data generating process is both regime-wise stationary and inconsistent with PPP, there is an ambiguity in the concepts of size and power when we evaluate the ADF tests and the PPP restricted structural change tests. Tests with good power to reject the unit root null will be oversized with regard to the PPP hypothesis. Since we are primarily concerned with PPP, we will focus on the latter. Both the ADF tests and the PPP restricted structural change tests generally have good size with medium (0.3) and large (0.5) breaks, being even undersized for all but very low levels of persistence. However, in the case of very small breaks (0.1) the results show substantial size distortion for all values of α. This is not surprising. As the size of the breaks decreases, the limit of a regime-wise stationary process will be a stationary process without breaks.

Next we consider a process which is consistent with PPP (the breaks are equal and of opposite sign). The ADF test has good power with small breaks at all but very high levels of persistence, moderate power for medium breaks and low levels of persistence (α= 0.5, α = 0.6), and low power for medium or large breaks with high levels of persistence (α= 0.7, α = 0.8). The QPPP test with two structural changes has good power in all cases, with the power decreasing as the process becomes more persistent. The PPP restricted structural change test has very high power when the process has breaks of equal and opposite sign, regardless of the magnitude of the break, except for the most highly persistent processes (α = 0.8). The simulation results show evidence of non-monotonic power for both the QPPP and the PPP restricted structural change tests at higher levels of persistence (α = 0.7, 0.8). As above, the power first decreases and then increases as the size of the break rises.

Generating trend stationary data which allows for two changes in the intercept and applying the previous tests, including a time trend, we obtain some similar and some fairly different results than in the case of non-trending generated data (Table 7). In the case of breaks that occur in the same direction, the ADF test and the TPPP restricted test have good power for all the break sizes, except for highly persistent data, because the time trend adjusts to compensate for the structural changes. The TPPP restricted structural change test has the highest power when

[21] After reaching a certain size, any additional increase in the magnitude of the break will not modify the power of the test.

[22] The issue of non-monotonic power in models with mean shifts or trend shifts is discussed by Vogelsang (1997, 1999).

the data generating process includes two structural changes that are equal and of opposite sign. The ADF test, in contrast, has much lower power on processes that are consistent with PPP (breaks that are equal and of opposite sign).

Finally, we conduct simulations to determine whether our tests can discriminate between PPP and TPPP restricted structural change. As reported in Table 8, we run unit root tests that incorporate PPP restricted structural change on various types of *trending* data. If the data is trending with no breaks, tests that incorporate PPP restricted structural change have essentially zero power. If the data is trending with breaks of equal and opposite sign (TPPP restricted structural change), tests that incorporate PPP restricted structural change have low power in the case of low persistence ($\alpha = 0.5, \alpha = 0.6$) and virtually zero power in the case of higher persistence ($\alpha = 0.7, \alpha = 0.8$). These results are in accord with the results of West (1987) on ADF tests that do not incorporate structural change.

We do not report the results of TPPP restricted structural change tests on various types of *non-trending* data because they are identical to the results reported earlier for *trending* data.[23] If the data is non-trending, either with no breaks or with breaks of equal and opposite sign (PPP restricted structural change), tests that incorporate TPPP restricted structural change have fairly good power, although not as good as tests that incorporate PPP restricted structural change. This is also in accord with previous work on ADF tests that do not incorporate structural change, where including an unnecessary time trend as a regressor will decrease power.

3.3 Interpretation of the empirical results

We proceed to interpret our empirical results in the context of the findings from the simulations. Recall that we rejected the unit root null at the 5% level for 9 out of 16 countries with the ADF test. Because the simulation evidence shows that the ADF tests have no power or very low power when the data contains any variant of structural change, this provides evidence of PPP or TPPP for these countries. Applying our new test, which both allows for structural change and imposes parity restrictions, we add 5 more rejections to the previous results: Denmark, Japan, Portugal, Switzerland and the United Kingdom. The combination of rejection of the unit root null with tests that incorporate restricted structural change and failure to reject the

[23] If a time trend is included as a regressor, it does not matter whether the data is generated with a zero or a non-zero trend.

unit root null with tests that do not incorporate restricted structural change provides evidence of PPP (TPPP) restricted structural change for these countries.

The simulation results also allow us to discriminate between evidence of PPP and TPPP restricted structural change. For Denmark, Japan and Switzerland the unit root null is rejected in favor of the restricted structural change alternative when the tests include a time trend, but not rejected when the tests do not include a time trend. This provides clear evidence of TPPP restricted structural change. For Portugal and the United Kingdom, the unit root null is rejected in favor of the restricted structural change alternative in both cases. Based on our simulation results, where tests for PPP restricted structural change have very low power when the data is actually generated by a model with TPPP restricted structural change, we interpret these findings as evidence of PPP restricted structural change.

As a further check on our findings of PPP and TPPP restricted structural change, we compare the break dates and coefficients from the restricted structural change models (Table 4) to those from the unrestricted two structural change models (Table 3). For Portugal and the United Kingdom, the comparison reinforces the findings of PPP restricted structural change. For Denmark, Japan and Switzerland it reinforces the findings of TPPP restricted structural change. The coefficients on the breaks in the unrestricted models are of opposite sign, and the break dates do not change dramatically with the imposition of the restrictions. These results are illustrated in Figure 1.

On the other hand, we did not find any variant of PPP in Canada and the Netherlands. In the case of Canada one possibility for the lack of evidence of PPP is the very large depreciation of the Canadian dollar against the US dollar in the end of the sample. The Netherlands experiences only one structural change, which is not consistent with either one of the PPP alternatives (Figure 2).

There are numerous political and economic factors that have the potential to cause shocks to real exchange rates. We explore possible explanations for some of the structural changes found with restricted models. Most of the countries where we found evidence of PPP or TPPP restricted structural change experienced a depreciation of their exchange rate against the dollar associated with the two World Wars (Portugal, the United Kingdom, Denmark and Switzerland). On the other hand, the collapse of the Bretton Woods system, in 1971, triggered a positive shock in the United Kingdom, Japan and Switzerland, causing a strong appreciation of their exchange rates against the dollar. Following the establishment of flexible nominal exchange rates, these countries experienced a return to the original level (or trend) of their real exchange rates.

4. Conclusion

Does long-run purchasing power parity hold between the United States and other industrialized countries? If there is reversion in the long run, is it to a constant mean, as in the version of PPP developed by Cassel, or is it to a constant trend, as in the version of PPP theory developed by Balassa and Samuelson. In order to make the distinction clear, we differentiate two concepts: Purchasing Power Parity (PPP) and Trend Purchasing Power Parity (TPPP).

Using the longest span of available historical data, we first perform conventional Augmented Dickey-Fuller (ADF) unit root tests which, with over a century of data, provide sufficient power to reject the unit root null in favor of a level (or trend) stationary alternative. We find evidence of PPP for 8 countries, Belgium, Germany, Finland, France, Italy, Norway, Spain, and Sweden, and evidence of TPPP for one additional country, Australia.

We proceed to investigate the hypothesis that the failure to reject unit roots in some of the real exchange rates can be explained by the presence of structural change. As a first step, we test for unit roots in the presence of a one time change in the intercept or a one time shift in the trend function. We find evidence of either Qualified PPP or Trend Qualified PPP for 11 out of 16 countries. Next we extend the previous tests to include two structural changes. In this case we find evidence of either QPPP or TQPPP for 12 out of 16 countries. These tests, however, do not provide evidence of either PPP or TPPP.

We then consider the possibility that a series may experience both structural change and reversion to its mean (or trend). We develop unit root tests that restrict the coefficients on the dummy variables that depict the breaks to produce a constant mean or trend in the long run. These restrictions ensure that the rejection of the unit root in favor of the PPP or TPPP restricted structural change is evidence of long-run (trend) purchasing power parity. With these new restricted tests, we add 5 countries to the previous PPP or TPPP evidence. Canada and the Netherlands are the only countries where we do not find evidence of any variant of PPP.

The simulation experiments reinforce our empirical results. There are three conclusions that we can draw: First, we find evidence that ADF tests have very low power to reject the unit root hypothesis in processes that incorporate structural change, including QPPP, TQPPP, and both PPP and TPPP restricted structural change. Rejections using ADF tests therefore provide strong evidence of PPP or TPPP without structural change. On the other hand, our restricted test has very good power when the process incorporates structural change that is consistent with the PPP or TPPP hypothesis, but low or moderate power in other cases. Finally, tests that incorporate

PPP restricted structural change have very low power if the data is generated by a model with TPPP restricted structural change, which allows us to discriminate between evidence of PPP and TPPP restricted structural change.

Taking in account the previous simulation results, we conclude that the restricted tests provide strong evidence of PPP restricted structural change for Portugal and the United Kingdom and TPPP restricted structural change for Denmark, Japan and Switzerland. This result is reinforced by comparing break dates and coefficients from unrestricted and restricted structural change models.

This paper posed two questions: Is there evidence of long-run PPP or TPPP among industrialized countries and, if so, which variant does the evidence support? Using conventional and restricted tests we find evidence of PPP and/or TPPP for 14 of the 16 countries. By including countries which experience structural change that is consistent with long-run PPP or TPPP, we increase the evidence by approximately 50% compared with conventional unit root tests. Using a combination of econometric and simulation evidence, we conclude that the Cassel version of PPP is supported for 10 countries and the Balassa-Samuelson version of TPPP is supported for 4 countries.

References

Abuaf, Niso and Philippe Jorion (1990), “Purchasing Power Parity in the Long Run,” *The Journal of Finance*, 45, 157-174.

Asea, Patrick K. and Max Corden (1994), “The Balassa-Samuelson Model: An Overview,” *Review of International Economics*, 2, 191-200.

Asea, Patrick K. and E. G. Mendoza (1994), “The Balassa-Samuelson Model: A General Equilibrium Appraisal,” *Review of International Economics*, 2, 244-267.

Balassa, Bela (1964) “The Purchasing Parity Power Doctrine: a Reappraisal,” *The Journal of Political Economy*, 72, 584-596.

Ben-David, Dan; Robin Lumsdaine and David H.Papell (2003), “Unit Roots, Postwar Slowdowns and Long Run Growth: Evidence from Two Structural Breaks,” *Empirical Economics,* 28, 303-319.

Cassel, Gustav (1918), “Abnormal deviations in international exchanges,” *The Economic Journal*, 28, 413-415.

Canzoneri, Matthew B., Robert E. Cumby and Behzad Diba (1999), “Relative Labor Productivity and the Real Exchange Rate in the Long Run: Evidence for a Panel of OECD Countries,” *Journal of International Economics*, 47, 245-266.

Cheung,Yin Wong and Kon S. Lai (2000), “On the Purchasing Parity Puzzle,” *Journal of International Economic,* 52, 321-330.

Cuddington, John T. and Hong Liang (2000), "Purchasing Power Parity over two centuries?", *Journal of International Money and Finance,* 19, 753-757.

Dornbusch, Rudiger (1988), “Purchasing Power Parity,” The *New Palgrave Dictionary of Economics*, Macmillan.

Dornbusch, Rudiger and Timothy Vogelsang (1991), “Real Exchange Rates and Purchasing Power Parity,” *Trade Theory and Economic Reform: North, South and East, Essays in Honnor of Bela Balassa*, Cambridge, MA, Basil Blackwell.

Elliott, Graham, Thomas J. Rothenberg and James H. Stock (1996), “Efficient Tests for an Autoregressive Unit Root,” *Econometrica*, 64, 813-836.

Engel, Charles (2000), “Long run PPP may not hold after all,” *Journal of International Economics*, 57, 243-273.

Fitzgerald, Doireann (2003), “Terms-of-Trade Effects, Interdependence and Cross-Country Differences in Price Levels”, working paper, Harvard University.

Hegwood, Natalie D. and David H. Papell (1998), "Quasi Purchasing Power Parity," *International Journal of Finance and Economics,* 3, 279-289.

Lumsdaine, Robin L. and David H. Papell (1997), "Multiple Trend Breaks and the Unit Root Hypothesis," *The Review of Economics and Statistics*, May 1997, 212-218.

Lopez, Claude, Christian J. Murray and David H. Papell (2002), "State of the Unit Root Tests and the PPP Puzzle," working paper, University of Houston.

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 (2000), "Purchasing Power Parity Over Two Centuries: Strengthening the Case for Real Exchange Rate Stability: A Reply to Cuddington and Liang," *Journal of International Money and Finance*, 19, 759-764.

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

Ng, Serena and Pierre Perron (2001), "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power," *Econometrica*, 69, 1519-1554.

Ng, Serena and Pierre Perron (2002), "PPP May not Hold After All: a Further Investigation," *Annals of Economics and Finance,* 3, 41-64.

O'Connell, Paul G.J. (1998), "The Overvaluation of Purchasing Power Parity," *Journal of International Economics*, 44, 1-19.

Obstfeld, Maurice (1993), "Model Trending Real Exchange Rates," Center for International and Development Economic Research, working paper no. C93-011.

Papell, David H. (2002), "The Great Appreciation, the Great Depreciation and the Purchasing Power Parity Hypothesis," *Journal of International Economics*, 57, 51-82.

Papell, David H. and Hristos Theodoridis (2001), "The Choice of Numeraire Currency in Panel Tests of Purchasing Power Parity," *Journal of Money, Credit and Banking*, August, 790-803.

Perron, Pierre (1997), "Further Evidence on Breaking Trend Functions in Macroeconomic Variables," *Journal of Econometrics,* 80, 355 – 385.

Perron, Pierre and Timothy J. Voselgang (1992), "Nonstationarity and Level Shifts with an Application to Purchasing Power Parity," *Journal of Business and Economic Statistics*, 10, 301-320.

Rogoff, Kenneth (1996), "The Purchasing Power Parity Puzzle," *Journal of Economic Literature*, 34, 647-668.

Samuelson, Paul A. (1994), "Facets of Balassa-Samuelson Thirty Years Later," *Review of International Economics*, 2, 201-226.

Samuelson, Paul A. (1964) “Theoretical Notes on Trade Problems,” *The Review of Economics and Statistics*, 46, 145-154.

Taylor, Alan (2002), “A Century of Purchasing-Power Parity,” *Review of Economics and Statistics,* 84, 139-150.

Vogelsang, Timothy J. and Pierre Perron (1998), “Additional Tests for a Unit Root Allowing for a Break in the Trend Function at an Unknown Time,” *International Economic Review*, 39, 1073-1100.

Vogelsang, Timothy J. (1997), “Sources of nonmonotonic power when testing for a shift in the trend of a dynamic time series,” *Center for Analytic Economics,* Working Paper #97-4*,* Cornell University.

Vogelsang, Timothy J. (1999), “Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series,” *Journal of Econometrics,* 88, 283-299.

West, Kenneth (1987), “A Note on the Power of Least Squares Tests for a Unit Root,” *Economics Letters*, 24, 249-252.

Table 1. Unit root tests

Augmented-Dickey-Fuller test: $\Delta y_t = \mu + \alpha y_{t-1} + \sum_{i=1}^{k} c_i \Delta y_{t-i} + \varepsilon_t$

T	Real exchange rate	α	k	t_α
129	Australia	-0.10261	1	-2.62*
117	Belgium	-0.22246	1	-4.12 ***
129	Canada	-0.06993	0	-1.62
119	Denmark	-0.06547	6	-1.24
119	Germany	-0.09041	1	-2.95 **
118	Finland	-0.41589	1	-6.02 ***
119	France	-0.15629	1	-3.55***
119	Italy	-0.24655	2	-4.28 ***
114	Japan	-0.01636	1	-1.02
129	Netherlands	-0.09555	1	-2.79 *
129	Norway	-0.12918	1	-3.67 ***
109	Portugal	-0.11736	5	-2.25
119	Spain	-0.12525	1	-3.24**
119	Sweden	-0.17148	1	-3.72 ***
107	Switzerland	-0.04288	2	-1.5
129	United Kingdom	-0.14768	4	-2.61 *

*, **, *** denote significance at the 10%, 5% and 1% level of significance, respectively.
The critical values for t_α are: -2.58 (10%), -2.89 (5%) and -3.51 (1%).

Augmented-Dickey-Fuller test with trend: $\Delta y_t = \mu + \beta t + \alpha y_{t-1} + \sum_{i=1}^{k} c_i \Delta y_{t-i} + \varepsilon_t$

T	Real exchange rate	α	k	t_α
129	Australia	-0.17342	1	-3.66 **
117	Belgium	-0.30877	1	-5.05 ***
129	Canada	-0.14277	0	-2.98
119	Denmark	-0.12240	6	-1.97
119	Germany	-0.11217	1	-3.32*
118	Finland	-0.44080	1	-6.21 ***
119	France	-0.22107	1	-4.16***
119	Italy	-0.24778	2	-4.27***
114	Japan	-0.08050	7	-1.98
129	Netherlands	-0.11675	1	-3.2 *
129	Norway	-0.15054	1	-3.96 **
109	Portugal	-0.12920	5	-2.15
119	Spain	-0.12933	1	-3.23 *
119	Sweden	-0.25125	1	-4.52 ***
107	Switzerland	-0.12977	2	-2.78
129	United Kingdom	-0.17255	4	-2.74

*, **, *** denote significance at the 10%, 5% and 1% level of significance, respectively.
The critical values for t_α are: -3.15 (10%), -3.45 (5%) and -4.04 (1%).

Table 2. Unit root tests including one structural change

QPPP test including one structural change

Real exchange rate	α	Break	γ	k	t_α
Australia	-0.238	1913	-0.26	1	-4.75**
Belgium	-0.312	1937	0.35	1	-5.12***
Canada	-0.197	1974	-0.13	1	-3.68
Denmark	-0.359	1968	0.38	3	-4.91**
Germany	-0.127	1984	0.27	5	-3.7
Finland	-0.502	1916	0.09	1	-9.85***
France	-0.327	1914	-0.27	1	-5.53***
Italy	-0.290	1941	0.10	2	-5.22***
Japan	-0.105	1962	1.01	7	-2.73
Netherlands	-0.214	1970	0.38	1	-4.56**
Norway	-0.204	1967	0.31	1	-4.85**
Portugal	-0.237	1920	-0.41	1	-4.19*
Spain	-0.185	1916	-0.18	1	-4.69**
Sweden	-0.254	1961	0.22	1	-4.63**
Switzerland	-0.240	1970	0.62	1	-5.03**
United Kingdom	-0.237	1941	-0.14	1	-4.11

*, **, *** denote significance at the 10%, 5% and 1% level of significance, respectively.
The critical values for t_α are: -4.20 (10%), -4.46 (5%) and -5.05 (1%).

TQPPP test including one structural change

Real exchange rate	α	Break	γ	k	t_α
Australia	-0.237	1913	-0.20	1	-4.7*
Belgium	-0.446	1920	-0.59	1	-6.2***
Canada	-0.387	1895	0.12	7	-4.26
Denmark	-0.369	1968	0.42	3	-4.99*
Germany	-0.147	1943	-0.30	1	-4.07
Finland	-0.535	1916	0.01	1	-10.83***
France	-0.432	1970	0.32	1	-6.16***
Italy	-0.292	1941	0.29	2	-5.03**
Japan	-0.238	1927	-0.70	1	-4.88*
Netherlands	-0.288	1962	0.38	7	-4.33
Norway	-0.195	1915	-0.11	1	-5.37**
Portugal	-0.280	1912	-0.41	5	-4.18
Spain	-0.192	1916	-0.23	1	-4.88*
Sweden	-0.298	1916	0.02	1	-5.38**
Switzerland	-0.280	1970	0.41	1	-5.31**
United Kingdom	-0.291	1943	-0.28	1	-4.83*

*, **, *** denote significance at the 10%, 5% and 1% level of significance, respectively.
The critical values for t_α are: -4.72 (10%), -5.02 (5%) and -5.61 (1%).

Table 3. Unit root tests including two structural changes

QPPP test including two structural changes

Real exchange rate	α	Break 1	γ_1	Break 2	γ_2	k	t_α
Australia	-0.279	1913	-0.20	1947	-0.11	1	-5.63**
Belgium	-0.360	1921	-0.33	1937	0.59	1	-5.51**
Canada	-0.317	1912	-0.03	1983	-0.15	4	-4.38
Denmark	-0.509	1939	-0.07	1967	0.42	4	-6.03***
Germany	-0.163	1930	0.43	1943	-0.25	1	-4.33
Finland	-0.566	1916	0.05	1970	0.12	1	-11.37***
France	-0.383	1914	-0.28	1982	0.05	1	-6.58***
Italy	-0.374	1919	-0.28	1942	0.27	2	-7.11***
Japan	-0.116	1930	0.23	1968	0.96	1	-3.68
Netherlands	-0.310	1936	-0.08	1965	0.39	4	-5.36*
Norway	-0.258	1915	0.01	1968	0.32	1	-6.32***
Portugal	-0.348	1916	-0.45	1986	0.34	1	-6.06***
Spain	-0.241	1916	-0.23	1982	0.25	1	-5.86**
Sweden	-0.329	1916	0.11	1963	0.18	1	-5.54**
Switzerland	-0.302	1930	0.18	1970	0.53	1	-5.68**
United Kingdom	-0.610	1944	-0.25	1972	0.19	3	-7.19***

*, **, *** denote significance at the 10%, 5% and 1% level of significance, respectively.
The critical values for t_α are: -4.89 (10%), -5.48 (5%) and -6.03 (1%).

TQPPP test including two structural changes

Real exchange rate	α	Break 1	γ_1	Break 2	γ_2	k	t_α
Australia	-0.295	1914	-0.30	1947	-0.20	1	-5.48
Belgium	-1.160	1909	-0.04	1927	-0.23	8	-9.14***
Canada	-0.532	1899	0.11	1951	-0.07	7	-4.91
Denmark	-0.556	1940	-0.13	1967	0.40	4	-6.13**
Germany	-0.188	1930	0.30	1943	-0.38	1	-4.79
Finland	-0.534	1916	-0.03	1947	-0.14	1	-11.71***
France	-0.484	1970	0.30	1982	0.05	1	-7.6***
Italy	-0.394	1919	-0.32	1942	0.22	2	-7.53***
Japan	-0.323	1927	-0.51	1972	0.36	1	-5.83*
Netherlands	-0.320	1936	0.20	1965	0.25	4	-5.37
Norway	-0.298	1914	0.18	1969	0.45	2	-6.58***
Portugal	-0.550	1916	-0.66	1948	-0.53	2	-7.13***
Spain	-0.281	1916	-0.42	1948	-0.52	1	-6.70***
Sweden	-0.332	1916	-0.05	1948	-0.24	1	-6.02**
Switzerland	-0.360	1943	-0.24	1971	0.38	1	-6.10**
United Kingdom	-0.623	1944	-0.29	1972	0.16	3	-7.32***

*, **, *** denote significance at the 10%, 5% and 1% level of significance, respectively.
The critical values for t_α are: -5.63 (10%), -5.91 (5%) and -6.41 (1%).

Table 4. Restricted structural change tests

PPP restricted structural change test

Real exchange rate	α	Break 1	γ_1	Break 2	γ_2	k	t_α
Australia	-0.198	1882	0.26	1913	-0.26	1	-4.10
Belgium	-0.292	1922	-0.59	1930	0.59	1	-4.85*
Canada	-0.133	1884	0.08	1948	-0.08	1	-2.61
Denmark	-0.144	1891	-0.20	1982	0.20	1	-3.16
Germany	-0.154	1911	-0.01	1980	0.01	8	-4.04
Finland	-0.507	1916	-0.03	1974	0.03	1	-10.24***
France	-0.336	1914	-0.21	1982	0.21	1	-6.42***
Italy	-0.372	1919	-0.28	1942	0.28	2	-7.06***
Japan	-0.060	1962	0.59	1985	-0.59	7	-2.42
Netherlands	-0.196	1882	-0.31	1968	0.31	1	-4.34
Norway	-0.173	1915	-0.12	1971	0.12	1	-5.05**
Portugal	-0.343	1916	-0.41	1986	0.41	1	-6.04***
Spain	-0.241	1916	-0.23	1968	0.23	1	-5.85***
Sweden	-0.230	1968	0.24	1977	-0.24	1	-4.67*
Switzerland	-0.089	1968	0.38	1987	-0.38	1	-2.86
United Kingdom	-0.578	1944	-0.23	1972	0.23	3	-7.08***

*, **, *** denote significance at the 10%, 5% and 1% level of significance, respectively.
The critical values for t_α are: -4.61 (10%), -4.93 (5%) and -5.48 (1%).

TPPP restricted structural change test

Real exchange rate	α	Break 1	γ_1	Break 2	γ_2	k	t_α
Australia	-0.242	1913	-0.04	1947	0.04	1	-5.40*
Belgium	-0.702	1910	-0.11	1971	0.11	7	-8.55***
Canada	-0.498	1895	0.08	1982	-0.08	8	-4.91
Denmark	-0.424	1944	-0.38	1966	0.38	3	-5.65**
Germany	-0.185	1930	0.33	1943	-0.33	1	-4.71
Finland	-0.560	1926	-0.05	1974	0.05	1	-11.60***
France	-0.388	1914	-0.16	1982	0.16	1	-6.63***
Italy	-0.372	1919	-0.27	1942	0.27	2	-7.06***
Japan	-0.313	1927	-0.43	1972	0.43	1	-5.83**
Netherlands	-0.230	1944	-0.36	1968	0.36	1	-4.88
Norway	-0.238	1915	-0.15	1967	0.15	1	-6.24***
Portugal	-0.345	1916	-0.38	1986	0.38	1	-5.92**
Spain	-0.242	1916	-0.22	1968	0.22	1	-5.74**
Sweden	-0.327	1916	0.00	1977	-0.00	1	-5.87**
Switzerland	-0.349	1943	-0.31	1970	0.31	1	-6.12***
United Kingdom	-0.588	1944	-0.23	1972	0.23	3	-7.07***

*, **, *** denote significance at the 10%, 5% and 1% level of significance, respectively.
The critical values for t_α are: -5.34 (10%), -5.59 (5%) and -6.12 (1%).

Table 5. Power against no structural change

Non-trending data generating processes

	ADF test			**QPPP – two structural change test**			**PPP restricted structural change test**		
	1% (-3.51)	5% (-2.89)	10% (-2.58)	1% (-5.59)	5% (-5.34)	10% (-4.89)	1% (-5.48)	5% (-4.93)	10% (-4.61)
1. Stationary generated data: $q_t = \mu + \alpha q_{t-1} + \varepsilon_t$									
$\alpha = 0.7$	68.2	88.8	94.7	58.1	87.3	94.3	80.8	96.4	99.2
$\alpha = 0.8$	47.6	77.6	88	21.1	51.2	69	42.4	75.3	88
$\alpha = 0.9$	13.8	42.4	58.2	3.6	18.1	31.8	7.2	29.3	46.6

Trending data generating processes

	ADF test			**TQPPP – two structural change test**			**TPPP restricted structural change test**		
Size and critical values	1% (-4.04)	5% (-3.45)	10% (-3.15)	1% (-6.41)	5% (-5.91)	10% (-5.63)	1% (-6.12)	5% (-5.59)	10% (-5.34)
2. Trend-stationary generated data: $q_t = \mu + \beta t + \alpha q_{t-1} + \varepsilon_t$									
$\alpha = 0.7$	55.5	79.8	88.8	48.5	76.9	88.8	53	83.5	91.9
$\alpha = 0.8$	29.2	61.1	75.5	16.3	42.4	59.8	17.9	45.3	63.2
$\alpha = 0.9$	8	24.9	41.7	3.5	12.5	25	3	14.9	25.1

Table 6. Power against two structural changes – non-trending data generating process

	ADF test			QPPP – two structural change test			PPP restricted structural change test		
	1% (-3.51)	5% (-2.89)	10% (-2.58)	1% (-6.03)	5% (-5.48)	10% (-5.21)	1% (-5.48)	5% (-4.93)	10% (-4.61)
3. Stationary generated data with two breaks in the intercept: $q_t = \mu + \alpha q_{t-1} + \gamma_1 DU1_t + \gamma_2 DU2_t + \varepsilon_t$									
I. $\alpha = 0.5$ **a) Coefficients on the breaks are equal and have the same sign**									
$\gamma_1 = 0.1, \gamma_2 = 0.1$	57.9	80.9	88.6	97.4	99.4	99.6	93.5	97.1	98.2
$\gamma_1 = 0.3, \gamma_2 = 0.3$	0.7	21.7	48.4	94.3	98.3	98.9	5.2	17	30.2
$\gamma_1 = 0.5, \gamma_2 = 0.5$	0	0	0.2	93.3	98	98.6	0	0	0.1
b) Coefficients on the breaks are equal and have opposite signs									
$\gamma_1 = 0.1, \gamma_2 = -0.1$	76.8	92.4	97.4	97.2	99.5	99.6	98.9	99.3	99.6
$\gamma_1 = 0.3, \gamma_2 = -0.3$	39.3	69.4	82.8	94.1	98.4	99	97.7	98.8	99.6
$\gamma_1 = 0.5, \gamma_2 = -0.5$	3.5	34.8	57.9	92.6	97.6	98.2	96.6	98.6	99.6

II. $\alpha = 0.6$ **a) Coefficients on the breaks are equal and have the same sign**									
$\gamma_1 = 0.1, \gamma_2 = 0.1$	43.4	71.3	82.9	87.8	97.4	99	80	91.6	96.3
$\gamma_1 = 0.3, \gamma_2 = 0.3$	0	1.8	9.5	80.1	94.4	97.4	4	2.7	7.7
$\gamma_1 = 0.5, \gamma_2 = 0.5$	0	0	0	80.7	94.4	97.4	0	0	0
b) Coefficients on the breaks are equal and have opposite signs									
$\gamma_1 = 0.1, \gamma_2 = -0.1$	68.3	88	94.8	88	97.8	99	96.3	99.2	99.5
$\gamma_1 = 0.3, \gamma_2 = -0.3$	16.4	50.6	68.3	78.5	94.8	97.7	91.9	97.6	99.3
$\gamma_1 = 0.5, \gamma_2 = -0.5$	0.5	8.3	24.2	81	94.7	97.1	93.3	97.5	99.3

Table 6. continued

	ADF test			QPPP – two structural change test			PPP restricted structural change test		
	1% (-3.51)	5% (-2.89)	10% (-2.58)	1% (-6.03)	5% (-5.48)	10% (-4.89)	1% (-5.48)	5% (-4.93)	10% (-4.61)
4. Stationary generated data with two breaks in the intercept: $q_t = \mu + \alpha q_{t-1} + \gamma_1 DU1_t + \gamma_2 DU2_t + \varepsilon_t$									
I. $\alpha = 0.7$ **a) Coefficients on the breaks are equal and have the same sign**									
$\gamma_1 = 0.1, \gamma_2 = 0.1$	24.2	52.3	69.5	56.4	85.6	93.3	40.8	70.6	82.9
$\gamma_1 = 0.3, \gamma_2 = 0.3$	0	0.1	0.5	43.7	76.9	88.8	0	0.1	0.6
$\gamma_1 = 0.5, \gamma_2 = 0.5$	0	0	0	54.2	81.3	90.1	0	0	0
b) Coefficients on the breaks are equal and have opposite signs									
$\gamma_1 = 0.1, \gamma_2 = -0.1$	53.9	80.8	89	57.7	85.5	93	78.8	95.7	98.4
$\gamma_1 = 0.3, \gamma_2 = -0.3$	4.9	22.4	42.1	43.1	76	88.2	69.5	91.9	96.4
$\gamma_1 = 0.5, \gamma_2 = -0.5$	0	0.8	4	54.5	82.9	91.6	79.7	95.2	97.7

II. $\alpha = 0.8$ **c) Coefficients on the breaks are equal and have the same sign**									
$\gamma_1 = 0.1, \gamma_2 = 0.1$	8	24.6	39.1	19	48.8	66.8	8.1	23.6	40
$\gamma_1 = 0.3, \gamma_2 = 0.3$	0	0	0	12.3	40.1	55.5	0	0	0
$\gamma_1 = 0.5, \gamma_2 = 0.5$	0	0	0	25.9	54.6	69.8	0	0	0
d) Coefficients on the breaks are equal and have opposite signs									
$\gamma_1 = 0.1, \gamma_2 = -0.1$	30.3	62.6	76.9	18.5	49.3	66.3	36.1	71.4	85.8
$\gamma_1 = 0.3, \gamma_2 = -0.3$	0.3	5.3	12.5	12.6	39.6	55.5	33.1	64.6	81.3
$\gamma_1 = 0.5, \gamma_2 = -0.5$	0	0	0.2	26.7	57.9	72.2	54.9	83	92.5

Table 7. Power against two structural changes – trending data generating process

	ADF test			TQPPP – two structural change test			TPPP restricted structural change test		
	1% (-4.04)	5% (-3.45)	10% (-3.15)	1% (-6.41)	5% (-5.91)	10% (-5.63)	1% (-6.12)	5% (-5.59)	10% (-5.34)
5. Trend-stationary generated data with two breaks in the intercept: $q_t = \mu + \beta t + \alpha q_{t-1} + \gamma_1 DU1_t + \gamma_2 DU2_t + \varepsilon_t$									
I. $\alpha = 0.5$ **a) Coefficients on the breaks are equal and have the same sign**									
$\gamma_1 = 0.1, \gamma_2 = 0.1$	72.7	89	94.1	95.6	98.7	99.5	96.3	99.1	99.4
$\gamma_1 = 0.3, \gamma_2 = 0.3$	55.9	78.7	87.2	91.9	97	98.3	91.6	96.5	97.9
$\gamma_1 = 0.5, \gamma_2 = 0.5$	31.8	63.4	77.8	88.7	95.6	97.6	71.4	89.4	94.1
b) Coefficients on the breaks are equal and have opposite signs									
$\gamma_1 = 0.1, \gamma_2 = -0.1$	60.3	81.4	88.5	96.5	98.7	99.6	97	99.1	99.4
$\gamma_1 = 0.3, \gamma_2 = -0.3$	15	42.1	58	93.8	97.8	99	93.1	97.9	98.7
$\gamma_1 = 0.5, \gamma_2 = -0.5$	0.4	4.3	15.8	91.1	97.1	98.6	92	97.4	98.1

II. $\alpha = 0.6$ **e) Coefficients on the breaks are equal and have the same sign**									
$\gamma_1 = 0.1, \gamma_2 = 0.1$	65	82.9	90.7	80.9	94.4	97.7	85.8	95.9	97.9
$\gamma_1 = 0.3, \gamma_2 = 0.3$	43.4	70.6	81.7	69	90	95.5	66.3	89.5	93.8
$\gamma_1 = 0.5, \gamma_2 = 0.5$	21.4	49.8	66.3	63.2	86.5	92.5	28.2	62.7	76.8
f) Coefficients on the breaks are equal and have opposite signs									
$\gamma_1 = 0.1, \gamma_2 = -0.1$	53.1	74	83.9	82.2	95.2	98.1	84.9	96.8	98.7
$\gamma_1 = 0.3, \gamma_2 = -0.3$	5.3	20.5	36.2	72.4	92	96.5	75.1	93.8	96.3
$\gamma_1 = 0.5, \gamma_2 = -0.5$	0.1	0.6	2.3	72.1	91.2	96.3	77.3	93.6	96.1

Table 7. continued

	ADF test			TQPPP – two structural change test			TPPP restricted structural change test		
	1% (-4.04)	5% (-3.45)	10% (-3.15)	1% (-6.41)	5% (-5.91)	10% (-5.63)	1% (-6.12)	5% (-5.59)	10% (-5.34)
6. Trend-stationary generated data with two breaks in the intercept: $q_t = \mu + \beta t + \alpha q_{t-1} + \gamma_1 DU1_t + \gamma_2 DU2_t + \varepsilon_t$									
I. $\alpha = 0.7$ **a) Coefficients on the breaks are equal and have the same sign**									
$\gamma_1 = 0.1, \gamma_2 = 0.1$	52.8	75	84.3	47.4	76.6	87.9	54.6	83.7	91.5
$\gamma_1 = 0.3, \gamma_2 = 0.3$	30.3	56.6	73.2	30.2	61.5	78	24.7	58.5	73.5
$\gamma_1 = 0.5, \gamma_2 = 0.5$	0.9	34	54.3	28.8	57.1	73.9	5.1	21.1	33.2
b) Coefficients on the breaks are equal and have opposite signs									
$\gamma_1 = 0.1, \gamma_2 = -0.1$	40.1	62	74	46.4	77.7	89.1	52.6	82.7	90.8
$\gamma_1 = 0.3, \gamma_2 = -0.3$	0.9	5.7	11.2	36.1	65.9	81.8	38.8	71.4	84.6
$\gamma_1 = 0.5, \gamma_2 = -0.5$	0	0	0.2	40.6	69	83.1	50.4	79.5	88.3

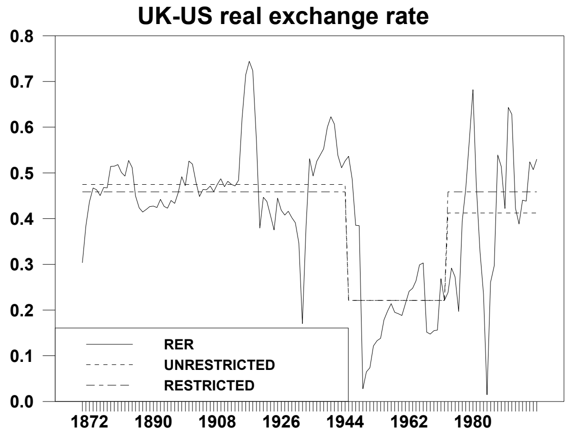
II. $\alpha = 0.8$ **g) Coefficients on the breaks are equal and have the same sign**									
$\gamma_1 = 0.1, \gamma_2 = 0.1$	31.3	59.5	72	13.3	40.3	58.8	15.2	46	63
$\gamma_1 = 0.3, \gamma_2 = 0.3$	13.3	37	54.4	6.6	23.3	40.4	4.8	18.7	31.7
$\gamma_1 = 0.5, \gamma_2 = 0.5$	1.6	13.6	32.5	5.9	23.1	37	0.6	3.7	7.5
h) Coefficients on the breaks are equal and have opposite signs									
$\gamma_1 = 0.1, \gamma_2 = -0.1$	17.7	39.7	55.1	15.2	40.2	60.2	15.5	43.4	60.9
$\gamma_1 = 0.3, \gamma_2 = -0.3$	0.2	0.4	1.9	7.6	28.6	45.6	10.4	34.5	48.4
$\gamma_1 = 0.5, \gamma_2 = -0.5$	0	0	0	14.8	37.6	54	23.9	50.9	66.6

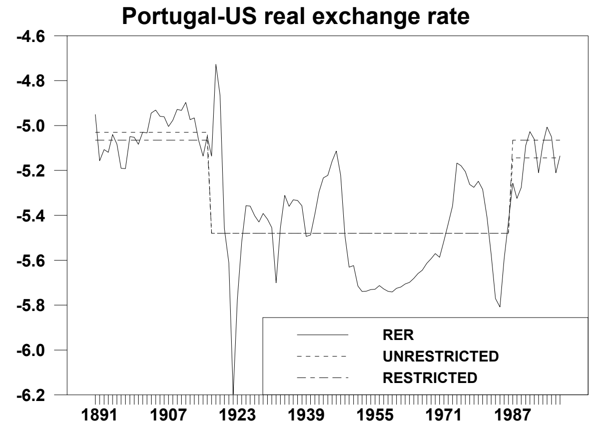
Table 8. Power of PPP restricted tests on trending data generating process

	PPP restricted structural change test		
	1% (-6.12)	5% (-5.59)	10% (-5.34)
Trend-stationary generated data: $q_t = \mu + \beta t + \alpha q_{t-1} + \varepsilon_t$			
α =0.7	0.2	1.1	3
α =0.8	0	0	0.2
α =0.9	0	0	0
Trend-stationary generated data with two breaks in the intercept: $q_t = \mu + \beta t + \alpha q_{t-1} + \gamma_1 DU1_t + \gamma_2 DU2_t + \varepsilon_t$			
α =0.5			
$\gamma_1 = 0.1, \gamma_2 = -0.1$	18.2	37.5	51
$\gamma_1 = 0.3, \gamma_2 = -0.3$	12.8	33.2	48.6
$\gamma_1 = 0.5, \gamma_2 = -0.5$	7.8	29.2	45.9
α =0.6			
$\gamma_1 = 0.1, \gamma_2 = -0.1$	2.5	10.3	22.3
$\gamma_1 = 0.3, \gamma_2 = -0.3$	1.9	8.2	18.7
$\gamma_1 = 0.5, \gamma_2 = -0.5$	1.5	8.5	19.6
α =0.7			
$\gamma_1 = 0.1, \gamma_2 = -0.1$	0.1	1	2.7
$\gamma_1 = 0.3, \gamma_2 = -0.3$	0	0.7	2.9
$\gamma_1 = 0.5, \gamma_2 = -0.5$	0.2	1.8	5.3
α =0.8			
$\gamma_1 = 0.1, \gamma_2 = -0.1$	0	0	0.2
$\gamma_1 = 0.3, \gamma_2 = -0.3$	0	0	0.5
$\gamma_1 = 0.5, \gamma_2 = -0.5$	0.1	0.5	1.8

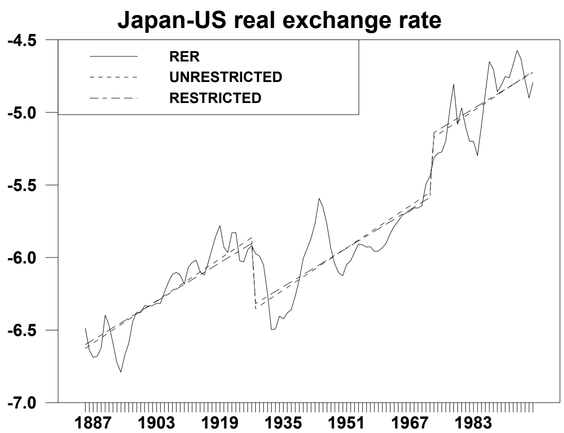
Figure 1. Evidence of PPP or TPPP restricted structural change

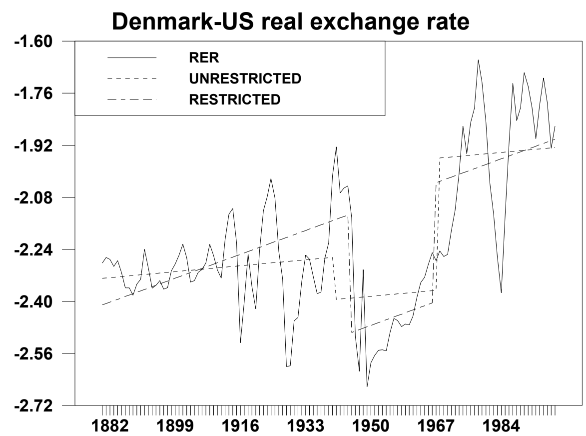
A. PPP restricted structural change





B. TPPP restricted structural change





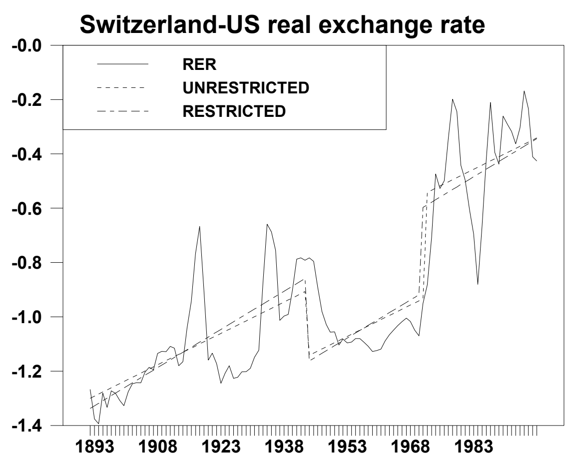


Figure 2. Countries with no evidence of PPP (TPPP)