Additional Evidence of Long Run Purchasing Power Parity with Restricted Structural Change

David H. Papell and Ruxandra Prodan

May 2005

We 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, using long-span real exchange rate data for industrialized countries. We develop unit root tests that both account for structural change and maintain a long run mean or trend. With conventional tests, previous research finds evidence of some variant of PPP for 9 of the 16 countries. With the unit root tests in the presence of restricted structural change, we find evidence of some variant of PPP for 5 additional countries.

We are grateful to participants in seminars at Berkeley, Davis, Emory, Michigan, Washington, the Federal Reserve Bank of San Francisco, Texas Camp Econometrics VIII, and the 2005 International Economics and Finance Society Meetings, as well as to an anonymous referee, for helpful comments and discussions. Papell thanks the National Science Foundation for financial support.

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. (205) 348-8985, email: rprodan@cba.ua.edu, Department of Economics, University of Alabama, Tuscaloosa, AL 35405.

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 goods 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-span real exchange rates. In the 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-span real exchange rates has been extensively investigated. 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. Taylor

[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.

(2002) develops a long-span nominal exchange rate and price level data set for 17 industrialized countries, producing 16 real exchange rates with the United States dollar as the numeraire currency. He finds that the unit root null hypothesis can be rejected at the 5% level in favor of either level or trend stationarity for 11 out of 16 real exchange rates using the Elliott, Rothenberg and Stock (1996) generalized-least-squares version of the Dickey-Fuller (DF-GLS) test with allowance for a deterministic trend.[4]

Lopez, Murray, and Papell (2005) argue that Taylor's conclusion is sensitive to his use of sub-optimal lag selection in unit root tests. Using standard lag selection methods, they find that ADF and DF-GLS tests produce the same result: the unit root null hypothesis can be rejected at the 5% level in favor of either level or trend stationarity for only 9 out of 16 real exchange rates of industrialized countries with the United States dollar as the numeraire currency. In this paper we analyze the 7 countries for which the unit root null cannot be rejected, proposing a new methodology in order to re-evaluate the Purchasing Power Parity hypothesis.

Much research has been conducted on testing for unit roots in the presence of a one-time change in the mean and/or 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 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.[5]

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

[4] Taylor also reports results with four Latin American countries and, for all countries, with a "world basket" as numeraire.

[5] 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).

time. As an extension, we propose tests that extend the previous models to incorporate two endogenous break points.

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. 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. Following this 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 either QPPP or TQPPP for 4 countries (at the 5% level).[6] These results are not necessarily a step forward towards PPP or TPPP. They do not impose either the alternative of a constant mean or the 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.[7]

We reject the unit root hypothesis in favor of PPP or trend PPP restricted structural change for 5 countries (at the 5% level). Combining these tests with the previous evidence from 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. In the 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 is a real appreciation of the guilder against the US dollar in the early 1970s following the discovery of large natural gas deposits that has not been reversed.

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

In order to interpret our findings, we conduct simulations to evaluate the power of the 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).[8] This implies that the previous 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.

Rogoff's celebrated (1996) "Purchasing Power Parity Puzzle" involves the combination of slow speeds of convergence to PPP and high short-term volatility of real exchange rates. Our paper can be considered to be a prelude to the PPP puzzle because evidence of PPP is necessary before one can sensibly measure the speed of convergence to PPP. Using tests that do not allow for structural change, Taylor (2002) found evidence of PPP or TPPP at the 5 percent level for 11 out of 16 industrial countries and Lopez, Murray, and Papell (2005) found evidence of PPP or TPPP for only 9 out of 16 countries. With the addition of tests that both allow for structural change and impose parity restrictions, evidence of PPP can be found for 10 countries and evidence of TPPP can be found for 4 additional countries, for a total of 14 out of 16 countries.

2. Testing for PPP and TPPP with 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 updated by the authors (using International Financial Statistics data). It consists of 107 to 129 years of real exchange rates for 16 industrialized countries with the United States dollar as the numeraire currency, starting between 1870 and 1892 and ending in 1998.

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

$$q_t = e_t + p_t * - p_t, \qquad (1)$$

where q_t is the logarithm of the real exchange rate, e_t 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.

[8] This result varies with the break size; the power of ADF test becomes higher if the size of the break is very small.

Previous research has provided evidence of PPP and TPPP using conventional unit root tests. Using the Elliott, Rothenberg and Stock (1996) generalized-least-squares version of the Dickey-Fuller (DF-GLS) test, Taylor (2002) rejects the unit root null in favor of either PPP or TPPP for 11 out of 16 industrialized countries with the U.S. dollar as numeraire at the 5% level and 4 additional countries at the 10% level. Using ADF tests, he finds evidence of either PPP or TPPP for 9 out of 16 countries at the 5% level and for 3 additional countries at the 10% level.

Lopez, Murray, and Papell (2005) argue that the reason for Taylor's strong rejections lies in the lag selection procedures used, which tend 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 ADF tests with general-to-specific lag selection, they 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, Belgium, Germany, Finland, France, Italy, Norway, Spain, and Sweden, 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.[9]

We proceed to analyze the 7 countries, Canada, Denmark, Japan, Netherlands, Portugal, Switzerland, United Kingdom and United States, where Lopez, Murray, and Papell (2005) could not reject the unit root null in favor of either level or trend stationarity.[10] First, we test for unit roots while allowing for structural change, but do not impose PPP or TPPP. Second, we test for unit roots in the presence of PPP or TPPP restricted structural change.

2.1. 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.[11]

---

[9] They report the same number of rejections using DF-GLS tests with MAIC lag selection.

[10] Given the results in Lothian and Taylor (1996), it may be surprising that we do not reject the unit root null for the United Kingdom. Hegwood and Papell (1998), however, show that, with general-to-specific lag selection, the unit root null is not rejected in favor of level stationarity, but is rejected in favor of stationarity around a one-time change in the mean, for the Lothian and Taylor dollar/sterling data.

[11] 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.

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. With long-span real exchange rates, most of the observations from nominal fixed exchange rate regimes, where devaluations and reevaluations, especially following failed attempts to defend currencies, can lead to discrete jumps (intercept changes).

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:

$$q_t = \mu + \gamma DU_t + \tilde{z}_t , \tag{2}$$

where DU_t = 1 if t > Tb and 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:

$$\Delta\tilde{z}_t = \sum_{i=0}^{k} \omega_i D(Tb)_{t-i} + \alpha\tilde{z}_{t-1} + \sum_{i=1}^{k} c_i \Delta\tilde{z}_{t-i} + \varepsilon_t , \tag{3}$$

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 $k\max$ = 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.[12]

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

$$q_t = \mu + \beta t + \gamma DU_t + \tilde{z}_t , \tag{4}$$

where 10% trimming is used to avoid finding spurious breaks. The unit root test is then performed using the t statistic for α = 0 in the regression described by Equation (3). 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 (3). The critical values for the finite sample distributions are taken from the sorted vector of 5000 replicated statistics.[13]

The results of the tests for a unit root in the presence of one structural change are reported in Table 1. We find evidence of Qualified Purchasing Power Parity (QPPP) by rejecting the unit root null for 3 out of 7 countries at the 5% level, Denmark, Netherlands 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 1 country at the 5% level, Switzerland, and for 3 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 3 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.[14] Following the previous testing procedures, the AO model is estimated using a two-step process. For values of the break points Tb_1 and Tb_2 with .10T< Tb_i <.90T (where T is the sample size and i = 1,2), the deterministic part of the series is removed using the following regression:

$$q_t = \mu + \gamma_1 DU1_t + \gamma_2 DU2_t + \tilde{z}_t \qquad (5)$$

where $DU1_t$ = 1 if t > Tb_1, 0 otherwise and $DU2_t$ = 1 if t > Tb_2, 0 otherwise. The unit root test is then performed using the t-statistic for α = 0 in the regression:

$$\Delta\tilde{z}_t = \sum_{i=0}^{k} \omega_{1i} D(Tb_1)_{t-i} + \sum_{i=0}^{k} \omega_{2i} D(Tb_2)_{t-i} + \alpha\tilde{z}_{t-1} + \sum_{i=1}^{k} c_i \Delta\tilde{z}_{t-i} + \varepsilon_t \,, \qquad (6)$$

where D(Tb_i)t = 1 if t = Tb_i +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:

$$q_t = \mu + \beta t + \gamma_1 DU1_t + \gamma_2 DU2_t + \tilde{z}_t \qquad (7)$$

---

[12] 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.
[13] We experimented by computing data specific critical values for several countries, and the results were unaffected.
[14] 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.

The unit root test is then performed using the t-statistic for α = 0 in the regression described by Equation (6). 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 2. We find evidence of Qualified Purchasing Power Parity (QPPP) by rejecting the unit root null for 4 out of 7 countries at the 5% level, Denmark, Portugal, 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 4 countries at the 5% level, Denmark, Portugal, 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 4 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 5 countries at the 5% level.

2.2. Tests for PPP in the presence of restricted structural change

Using an AO model and allowing for one or two intercept changes we found evidence of QPPP (or TQPP) for 5 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:

$$q_t = \mu + \gamma_1 DU1_t + \gamma_2 DU2_t + \tilde{z}_t , \qquad (8)$$

subject to the restriction: $\gamma_1 + \gamma_2 = 0$ (9)

where 10% trimming is used to avoid finding spurious breaks. The restriction in Equation (9) 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.

In order to make the unit root test invariant to $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$, we also impose the following restriction to Equation (6):

$$\omega_1 + \omega_2 = 0 \qquad (10)$$

The unit root test is then performed using the t statistic for α = 0 in the regression described by Equation (6). 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 (7), subject to the same restriction (9). 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 2. We find evidence of PPP Restricted Structural Change by rejecting the unit root null for 2 out of 7 countries at the 5% level, Portugal the United Kingdom. Incorporating time trends, we find evidence of Trend PPP Restricted Structural Change by rejecting the unit root null for 5 countries at the 5% level, Denmark, Japan, Portugal, Switzerland, and the United Kingdom. Combining the two tests, we find evidence of either PPP or TPPP Restricted Structural Change by rejecting the unit root null for 5 countries at the 5% level. Canada and Netherlands are the countries for which we do not find evidence of either PPP or TPPP Restricted Structural Change. In addition, we find evidence of QPPP for the Netherlands. Combined with previous evidence of Lopez, Murray, and Papell (2005), among the 16 industrialized countries, Canada is the only country for which there is no evidence of some variant of stationarity.

3. Power of the univariate unit root tests with structural change

By developing and implementing tests for a unit root in the presence of restricted structural change, we add 5 more rejections to unit roots in real exchange rates beyond those found by using conventional ADF or DF-GLS tests. 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 power of the newly developed tests and proceed to investigate their finite sample performance.

3.1 Construction of the power simulations

The simulation experiments address the following issues: (a) comparison of properties of the various tests (ADF, AO model allowing for two structural changes and AO restricted model), (b) power of AO models as a function of the magnitude of the break.

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), with two equal changes in the intercept of the same sign (12), and with two equal changes in the intercept of opposite sign (12). The first and third experiments are consistent with PPP or TPPP, the second is not. 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 = 0.8 and 0.9 for first generated sample (11) and α = 0.7 for the second generated sample (12). This reflects the range of values of α reported in Table 1. 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 (11) 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% (detailed results for different nominal sizes and values of α are presented in Table 3). First, as the errors become more persistent, the power decreases: The ADF test has good power when the data generating process is non-trending with α = 0.8 but lower power when α = 0.9. This is in accord with previous research, considering the span and persistence of the data. As expected, tests for QPPP that incorporate two structural changes have less power than the ADF test. Tests for PPP restricted structural change have generally lower power than the ADF test (except the case with lower persistence) but higher power than QPPP test. Table 3 also reports results for trending data generating processes. Applying the previous tests, including a time trend, the results are fairly similar. The power of the tests, however, is lower with trending than with non-trending data in most cases.

The results with stationary data generating processes (12) that allow for two changes in the intercept are reported in Table 4. 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 to reject the unit root null for all break sizes, although the simulation results show evidence of non-monotonic power: the power first decreases and then increases as the size of the break rises.[15]

Because the data generating process with two breaks that have the same sign is both regime-wise stationary and inconsistent with PPP, there is an ambiguity regarding how to interpret power. If

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

ADF and PPP restricted structural change tests have good power to reject the unit root null, they will correctly provide evidence of stationarity but incorrectly provide evidence of PPP. Since we are concerned with PPP, not stationarity, low power becomes a desirable property. Both the ADF tests and the PPP restricted structural change tests have extremely low power with medium (0.3) and large (0.5) breaks, but higher power in the case of very small breaks (0.1). Thus, at least for medium and large breaks, we are very unlikely to incorrectly find evidence of PPP. The result for very small breaks 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 low power for medium or large breaks. The QPPP test with two structural changes has good power in all cases. 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. The simulation results show evidence of non-monotonic power for both the QPPP and the PPP restricted structural change tests. 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 data (Table 4). In the case of breaks that occur in the same direction, the ADF test and the TPPP restricted test have a fairly good power for all the break sizes because the time trend adjusts to compensate for the structural changes. The TPPP restricted structural change test has the highest power when 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 TPPP (breaks that are equal and of opposite sign).

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 previous research 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 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. Due to the fact that 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. [16]

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 2) to those from the unrestricted two structural change models (Table 1). 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.

The speed of reversion to PPP has become an active research topic. Using the values of α in Table 2, we calculate, for our preferred specifications, the half-lives of PPP (or TPPP) deviations, the time that it takes for a shock to the real exchange rate to return halfway to its long run PPP (or TPPP) restricted mean (or trend). The half-lives are all under two years, ranging from .80 years for the United Kingdom to 1.97 years for Japan. The short half-lives should not be surprising because, by measuring the return to the restricted mean or trend, we have taken out the effects of two shocks that are of sufficient importance to cause the non-rejection of the unit root hypothesis in tests that do not incorporate structural change.[17]

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. This is the classic example of the "Dutch disease", the large real appreciation of the guilder following the discovery of large natural gas deposits in the North Sea.

What caused the structural changes to the real exchange rates? With the exception of the Netherlands, these changes either represent temporary (although long-lived) movements away from (T)PPP or movements that restore (T)PPP. In addition, the initial movement away from PPP is, in all cases, a real depreciation against the dollar, followed by a real appreciation to restore PPP. Three of the

[16] We conducted, but do not report, simulations that illustrate this result, which is in accord with the findings of West (1987) on ADF tests that do not incorporate structural change.

[17] We did not calculate the half-lives from the impulse response function or correct for median bias, as in Murray and Papell (2002) because the half-lives are so short that these corrections would not be particularly important.

countries where we found evidence of PPP or TPPP restricted structural change experienced a depreciation of their nominal exchange rates against the dollar following the end of World War I (Portugal) or World War II (Denmark and the United Kingdom). For Switzerland, while the nominal exchange rate remained fixed, U.S. prices rose faster than Swiss prices, causing a real depreciation starting in 1943. For Japan, the onset of the deviation from PPP was caused by a nominal depreciation in 1930. The offsetting shifts that restored PPP or TPPP are centered around the collapse of the Bretton Woods system of fixed exchange rates, which triggered a nominal appreciation against the dollar for Denmark, Japan and Switzerland, causing an appreciation of their real exchange rates. For the United Kingdom, the appreciation of the pound seems to be caused by high inflation relative to the United States. These movements in the nominal exchange rates and prices (ratios to US prices) are illustrated in Figure 2. 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 spirit of the version of PPP developed by Cassel, or is it to a constant trend, as in the spirit of 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).

Lopez, Murray, and Papell (2005) have previously found evidence of a variant of PPP for 9 out 16 industrialized countries. In this study we focus on the remaining 7 countries. We first 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 one or two changes in the intercept or shifts in the trend function. We find evidence of either Qualified PPP or Trend Qualified PPP for 4 out of 7 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. 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. Second, 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.

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. Most of the structural changes are associated with movements in nominal exchange rates.

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? Combining previous results of conventional tests with our new 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 5 countries compared with conventional unit root tests. Using a combination of econometric and simulation evidence, we conclude that PPP is supported for 10 countries and 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 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.

Campbell, John and Pierre Perron (1991), "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots," NBER Macroeconomics Annual, 141-201

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.

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.

Fitzgerald, Doireann (2003), "Terms-of-Trade Effects, Interdependence and Cross-Country Differences in Price Levels", working paper, University of California, Santa Cruz.

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*, LXXIX, 212-218.

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, 361-369.

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.

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

Ng, Serena and Pierre Perron, (1995), "Unit Root Tests in ARMA Models with Data Dependent Methods for the Selection of the Truncation Lag," *Journal of the American Statistical Association*, 90, 268-281.

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.

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.

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. (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 including one and two structural changes

Real exchange rate	α	Break	γ	k	t_α
QPPP test including one structural change					
Canada	-0.197	1974	-0.13	1	-3.68
Denmark	-0.359	1968	0.38	3	-4.91**
Japan	-0.105	1962	1.01	7	-2.73
Netherlands	-0.214	1970	0.38	1	-4.56**
Portugal	-0.237	1920	-0.41	1	-4.19*
Switzerland	-0.240	1970	0.62	1	-5.03**
United Kingdom	-0.237	1941	-0.14	1	-4.11
TQPPP test including one structural change					
Canada	-0.387	1895	0.12	7	-4.26
Denmark	-0.369	1968	0.42	3	-4.99*
Japan	-0.238	1927	-0.70	1	-4.88*
Netherlands	-0.288	1962	0.38	7	-4.33
Portugal	-0.280	1912	-0.41	5	-4.18
Switzerland	-0.280	1970	0.41	1	-5.31**
United Kingdom	-0.291	1943	-0.28	1	-4.83*

Real exchange rate	α	Break 1	γ_1	Break 2	γ_2	k	t_α
QPPP test including two structural changes							
Canada	-0.317	1912	-0.03	1983	-0.15	4	-4.38
Denmark	-0.509	1939	-0.07	1967	0.42	4	-6.03**
Japan	-0.116	1930	0.23	1968	0.96	1	-3.68
Netherlands	-0.310	1936	-0.08	1965	0.39	4	-5.36*
Portugal	-0.348	1916	-0.45	1986	0.34	1	-6.06***
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***
TQPPP test including two structural changes							
Canada	-0.532	1899	0.11	1951	-0.07	7	-4.91
Denmark	-0.556	1940	-0.13	1967	0.40	4	-6.13**
Japan	-0.323	1927	-0.51	1972	0.36	1	-5.83*
Netherlands	-0.320	1936	0.20	1965	0.25	4	-5.37
Portugal	-0.550	1916	-0.66	1948	-0.53	2	-7.13***
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:

-4.20 (10%), -4.46 (5%) and -5.05 (%) (QPPP test including one structural change)
-4.72 (10%), -5.02 (5%) and -5.61 (1%) (TQPPP test including one structural change)
-5.24 (10%), -5.51 (5%) and -6.06 (1%) (QPPP test including two structural changes)
-5.69 (10%), -5.96 (5%) and -6.45 (1%) (TQPPP test including two structural changes)

Table 2. Restricted structural change tests

Real exchange rate	α	Break 1	Break 2	γ_1 ($\gamma_1 = -\gamma_2$)	k	t_α
PPP restricted structural change						
Canada	-0.169	1886	1976	010	4	-2.72
Denmark	-0.121	1921	1948	-0.05	1	-2.98
Japan	-0.047	1944	1985	0.34	7	-2.27
Netherlands	-0.192	1882	1968	-0.31	1	-4.32
Portugal	-0.331	1916	1984	-0.39	1	-5.71***
Switzerland	-0.09	1967	1987	0.36	2	-2.86
United Kingdom	-0.578	1944	1972	-0.23	3	-7.20***
TPPP restricted structural change						
Canada	-0.469	1895	1983	0.08	7	-5.23
Denmark	-0.445	1944	1966	-0.36	3	-5.87**
Japan	-0.296	1930	1976	-0.42	1	-5.59**
Netherlands	-0.214	1916	1968	-0.25	1	-4.94
Portugal	-0.336	1916	1984	-0.36	1	-5.63**
Switzerland	-0.350	1943	1970	0.31	1	-6.22***
United Kingdom	-0.586	1944	1972	-0.23	3	-7.20***

*, **, *** denote significance at the 10%, 5% and 1% level of significance, respectively.
The critical values for t_α are:
-4.72 (10%), -5.04 (5%) and -5.67 (1%) (PPP restricted structural change)
-5.31 (10%), -5.59(5%) and -6.21 (1%) (TPPP restricted structural change)

Table 3. Power against no structural change

a) Non-trending data

	ADF test			**QPPP – two structural change test**			**PPP restricted structural change test**		
	1% (-3.57)	5% (-2.95)	10% (-2.64)	1% (-6.06)	5% (-5.51)	10% (-5.24)	1% (-5.67)	5% (-5.04)	10% (-4.72)
1. Stationary generated data: $q_t = \mu + \alpha q_{t-1} + \varepsilon_t$									
$\alpha = 0.8$	0.365	0.657	0.805	0.217	0.525	0.701	0.324	0.683	0.835
$\alpha = 0.9$	0.117	0.333	0.495	0.068	0.202	0.335	0.076	0.267	0.444

b) Trending data

	ADF test including a time trend			**TQPPP – two structural change test**			**TPPP restricted structural change test**		
	1% (-4.17)	5% (-3.57)	10% (-3.23)	1% (-6.45)	5% (-5.96)	10% (-5.69)	1% (-6.21)	5% (-5.59)	10% (-5.31)
2. Trend-stationary generated data: $q_t = \mu + \beta t + \alpha q_{t-1} + \varepsilon_t$									
$\alpha = 0.8$	0.214	0.461	0.628	0.174	0.424	0.597	0.171	0.470	0.643
$\alpha = 0.9$	0.063	0.203	0.334	0.049	0.164	0.267	0.045	0.166	0.274

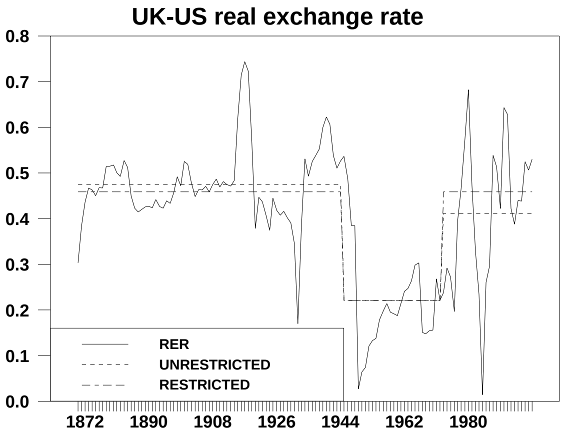
Table 4. Power against two structural changes – non-trending data generating process

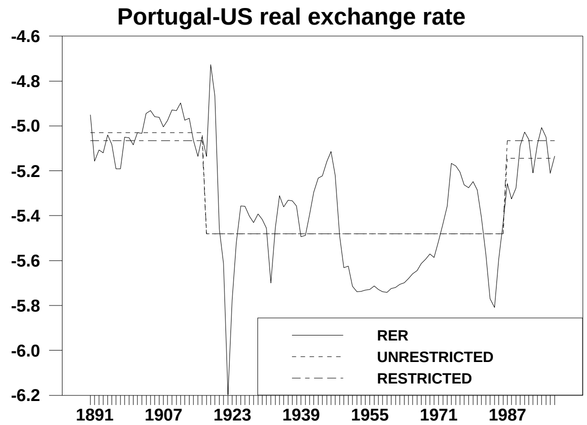
	ADF test			**QPPP – two structural change test**			**PPP restricted structural change test**		
$\alpha = 0.7$	1% (-3.57)	5% (-2.95)	10% (-2.64)	1% (-6.06)	5% (-5.51)	10% (-5.24)	1% (-5.67)	5% (-5.04)	10% (-4.72)
3. Stationary generated data with two breaks in the intercept: $q_t = \mu + \alpha q_{t-1} + \gamma_1 DU1_t + \gamma_2 DU2_t + \varepsilon_t$									
a) Coefficients on the breaks are equal and have the same sign									
$\gamma_1 = 0.1, \gamma_2 = 0.1$	0.211	0.455	0.615	0.525	0.814	0.907	0.297	0.621	0.784
$\gamma_1 = 0.3, \gamma_2 = 0.3$	0.000	0.000	0.002	0.398	0.702	0.837	0.000	0.001	0.001
$\gamma_1 = 0.5, \gamma_2 = 0.5$	0.000	0.000	0.000	0.476	0.755	0.865	0.000	0.000	0.000
b) Coefficients on the breaks are equal and have opposite signs									
$\gamma_1 = 0.1, \gamma_2 = -0.1$	0.489	0.756	0.780	0.533	0.829	0.918	0.676	0.913	0.962
$\gamma_1 = 0.3, \gamma_2 = -0.3$	0.033	0.177	0.377	0.381	0.710	0.833	0.554	0.879	0.945
$\gamma_1 = 0.5, \gamma_2 = -0.5$	0.000	0.000	0.011	0.453	0.769	0.863	0.664	0.906	0.960

	ADF test			**TQPPP – two structural change test**			**TPPP restricted structural change test**		
	1% (-4.17)	5% (-3.57)	10% (-3.23)	1% (-6.45)	5% (-5.96)	10% (-5.69)	1% (-6.21)	5% (-5.59)	10% (-5.31)
4. 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**									
$\gamma_1 = 0.1, \gamma_2 = 0.1$	0.528	0.750	0.843	0.454	0.725	0.853	0.461	0.783	0.886
$\gamma_1 = 0.3, \gamma_2 = 0.3$	0.303	0.566	0.732	0.251	0.567	0.730	0.223	0.494	0.654
$\gamma_1 = 0.5, \gamma_2 = 0.5$	0.009	0.340	0.543	0.233	0.561	0.697	0.037	0.168	0.287
b) Coefficients on the breaks are equal and have opposite signs									
$\gamma_1 = 0.1, \gamma_2 = -0.1$	0.401	0.620	0.740	0.432	0.717	0.872	0.464	0.769	0.884
$\gamma_1 = 0.3, \gamma_2 = -0.3$	0.009	0.057	0.112	0.334	0.622	0.767	0.345	0.646	0.788
$\gamma_1 = 0.5, \gamma_2 = -0.5$	0.000	0.000	0.002	0.395	0.666	0.795	0.402	0.708	0.818

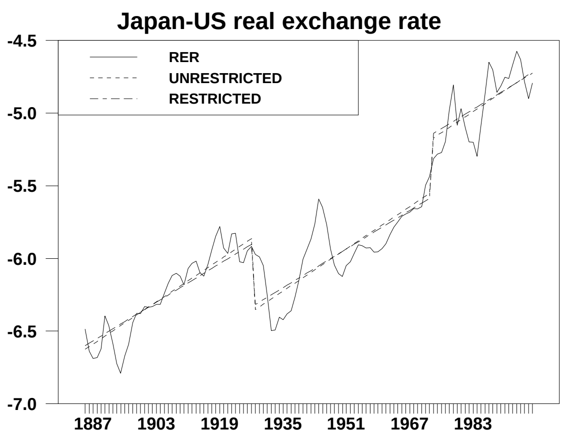
Figure 1. Evidence of PPP or TPPP restricted structural change

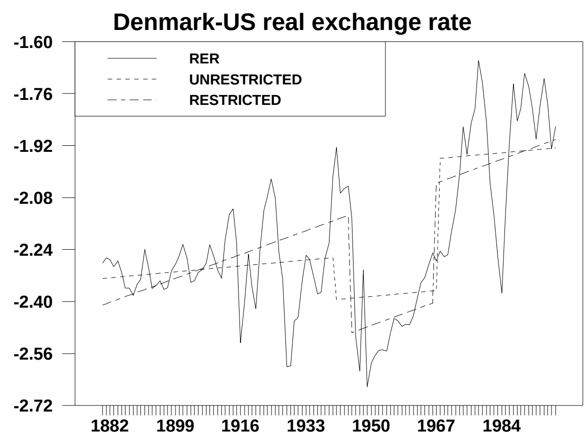
A. PPP restricted structural change





B. TPPP restricted structural change





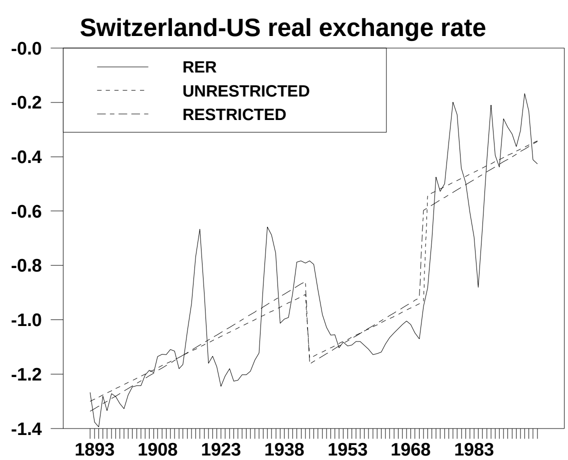
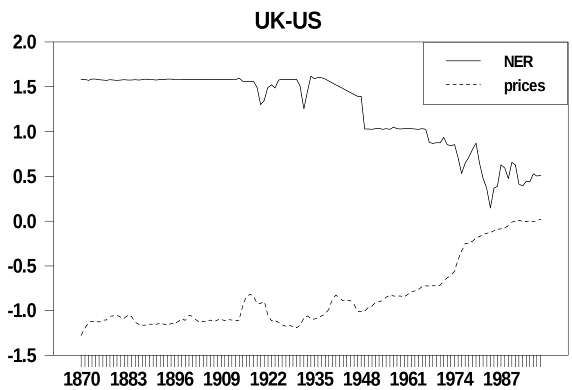
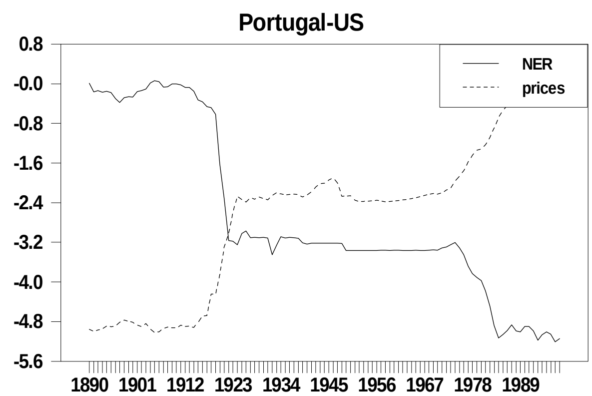


Figure 2. Nominal exchange rate and price (ratio to US) movements for countries where we found evidence of PPP or TPPP restricted structural change (in logs)

A. PPP restricted structural change





B. TPPP restricted structural change