Potential Pitfalls in Determining Multiple Structural Changes with an Application to Purchasing Power Parity

Ruxandra Prodan

University of Houston

November 2003

Tests for structural change play an important role in macroeconomics and international finance. We investigate the empirical performance of the Bai and Perron (1998) multiple structural change tests and show that the use of their critical values may cause severe size distortions in persistent series. To correct these size distortions, we implement the Bai (1999) structural change test, where we calculate bootstrap critical values. While the size is correct, his sequential method lacks power on data that includes breaks of opposite sign. We extend this test in two directions. First, we propose a new procedure to choose the number of breaks. Second, we develop a restricted version that specifically models breaks that imply mean or trend reversion. The simulation results show good power and satisfactory size of the new procedures. Using long-run real exchange rates, we illustrate the practical importance of these results. We investigate contradictory evidence of purchasing power parity from the application of unit root and structural change tests. We argue that the Bai and Perron test is the culprit and instead use the Bai test along with our newly proposed methodologies. Consequently, we reconcile the results of unit root tests with those of tests for structural change when testing for purchasing power parity.

I am grateful to Tim Hopper, Lutz Kilian, Ulrich Mueller, Chris Murray, Bent Sorensen, and especially David Papell, as well as to participants at the University of Houston Macroeconomics Workshop for helpful comments and discussions. I would also like to thank Pierre Perron for providing me with his Gauss program.

Correspondence to:

Department of Economics, University of Houston, Houston, TX 77204-5019, tel. (713) 743-3821, fax (713) 743-3798, email: rprodan@mail.uh.edu

Introduction

Structural change is of central importance to statistical modeling of time series. Disregarding this aspect could lead to inaccurate forecasts or inferences about economic relationships. Due to the importance of structural stability, much literature has been devoted to obtaining powerful tests for structural change in a variety of modeling contexts. The classic test for a structural change was developed by Chow (1960). His testing procedure, popular for many years, was extended to cover most econometric models of interest. Its important limitation, however, is that the break date must be known a priori.

In the early 1990's this problem was solved by several authors, the most significant studies being provided by Andrews (1993) and Andrews and Ploberger (1994).[1] They developed tests which use data-based procedures to determine the location of the break and derived their asymptotic distribution. All of these studies require either full specification of the dynamics or else consistent estimates of serial-correlation parameters, a poor handling of any of these leading to undesirable properties in finite samples. Vogelsang (1997) developed tests for detecting a break at an unknown date in the trend function of a dynamic univariate time series, allowing for trending and unit root regressors.

The development of tests for one structural change in the data led to the next question: could there be more than one? The problem of testing for multiple structural changes has received considerable attention. Bai and Perron (1998) consider estimating multiple structural changes, occurring at unknown dates, in a linear model estimated by least-squares. They also address the problem of testing for multiple structural changes, proposing a sequential procedure. First, search for the most significant break (if it exists). If a significant break is found, the sample is then split at the break point, and additional breaks can be found by searching over sub-samples. The procedure is continued until no additional significant breaks can be found. In a companion paper (2003a) they focus on the empirical implementation of their theoretical result, presenting an efficient algorithm, based on the principle of dynamic programming, which efficiently estimates models with multiple structural changes.

Bai and Perron (2001, 2003a and 2003b) provide new evidence on the adequacy of their methods and also identify several problems associated with these tests. While Bai and Perron (1998) recommend trimming, not allowing for breaks at the beginning and the end of the sample, of 5% of the data, they later argue that in the presence of serial correlation and/or heterogeneity in the data or errors across segments, a higher trimming is needed.[2] They also acknowledge the lack of power of the sequential procedure on data that includes certain configuration of changes (especially breaks of opposite sign), where it is difficult to

[1] Stock (1994) provides a useful survey.

[2] They argue that trimming of 5% can lead to substantial size distortions in these specific cases and advocate trimming of 15% or 20% to produce tests with correct size in finite samples.

reject the null hypothesis of 0 breaks versus the alternative hypothesis of 1 break, but it is not difficult to reject the null hypothesis of 0 breaks versus a higher number of breaks. To improve the power when the sequential procedure breaks down they propose what we call the *Bai and Perron procedure*: First, look at the sequential method. If 0 against 1 break is rejected, continue with the sequential method until the first failure to reject. If 0 against 1 is not rejected then look at the global tests to see if one break is present. If at least one break is found, decide the number of breaks based upon a sequential examination of $\ell+1$ against ℓ breaks with statistics constructed using estimates of the break dates obtained from a global minimization of the sum of squared residuals.

This study conducts a thorough analysis of the Bai and Perron multiple structural change tests. We provide new evidence that in persistent series the use of asymptotic critical values, calculated under the null of *iid* errors, can be inadequate. This causes serious problems with the interpretation of these tests in practice that are of relevance for many questions in macroeconomics and international finance. We first show that their sequential procedure may cause severe size distortions if the data are highly persistent. The existence of these size distortions has not been documented previously in the literature[3]. While Bai and Perron (2001) report some size results for the sequential method, their data generating process is an AR(1) with a coefficient of 0.5. They argue that the sequential procedure remains adequate when allowing for a lag of the dependent variable but shows some size distortion when using non-parametric methods to correct for serial correlation (an actual size of 8% for tests with a nominal size of 5%). In contrast, conducting a simulation experiment with persistent processes with AR coefficients between 0.5 and 0.99, we argue that the sequential method suffers from large size distortions (for a persistence of 0.9 we find a size of 20.1% and 21.4%, accounting for serial correlation in a parametric and non-parametric way, respectively).[4]

We proceed to analyze the *Bai and Perron procedure*. For persistent processes, this procedure, while improving the power on breaks of opposite sign, severely increases the evidence of size distortions compared with the sequential method. Even for white noise processes, the procedure shows some evidence of size distortions (a size of 10% on average with a nominal size of 5%). With a persistence of 0.9, using the *Bai and Perron procedure* we find a size of 35.2% and 43.3%, accounting for serial correlation in a parametric and non-parametric way, respectively, which represents a considerable increase from the sequential procedure.

Since it is known that many processes of interest in empirical macroeconomics and international finance tend to be highly persistent, our results of severe size distortions are of immediate practical

[3] Diebold and Chenn (1996) provide evidence of size distortions of supremum tests for a structural change in a dynamic model (Andrews, 1993 and Andrews and Ploberger, 1994).

[4] These values are calculated for a trimming of 15% while restricting the residuals to have the same variance. A lower trimming and/or a heterogeneous variance across residuals leads to higher size distortions.

interest. Among many examples, Rudebusch (1993), Diebold and Senhadji (1996), Lothian and Taylor (1996) and Taylor (2002) estimate models where the AR coefficients are 0.93, 0.82, 0.89 and 0.82 for quarterly postwar US real GNP, annual long-run US real GNP, the annual long-run dollar-sterling real exchange rate and 16 industrialized countries' (average) long-run dollar real exchange rates, respectively. Thus there is a good reason to think that an AR(1) with a coefficient of 0.8-0.9, rather than an AR(1) with a coefficient of 0.5, provides a better representation of many macroeconomic time series of interest.

We have identified an important problem related to the empirical performance of multiple structural change tests. We next propose a solution to this problem. To correct for the size distortions, we implement the Bai (1999) likelihood ratio type test. The Bai test is a sequential test where the breaks are being chosen globally. The essential feature of this test is its easy applicability which allows us to use straightforward methods in order to calculate bootstrap critical values. Using these methods, the test has an approximately correct size for any level of persistence. As in the case of the Bai and Perron test, however, the issue becomes the loss in power implied by the use of a sequential method. While the Bai test works fairly well on data that incorporates one structural change or two structural changes that occur in the same direction, it has low power on data that includes breaks of opposite sign.

In order to increase the power of the test on such configurations of changes, we extend the Bai test in two directions: First, in order to increase the power of the sequential method, we propose a similar methodology to the *Bai and Perron procedure* for situations where it is difficult to reject the hypothesis of 0 versus 1 but it is not difficult to reject the null hypothesis of 0 versus of a higher number of breaks, which we call the *modified Bai procedure*. Next, in order to further increase the power of this test when two or more changes that imply mean or trend reversion are present, we propose a restricted version of this test that specifically models these types of breaks (*the restricted procedure*). The simulation experiments show that by applying the *restricted procedure* we substantially increase the power performance of these tests on breaks of opposite sign, especially in the case of trending data, beyond the *modified Bai procedure*. However, as our simulation results show, while the power is improved, the size for any level of persistence becomes on average 10% (for tests with nominal size of 5%). While the sequential tests are correctly sized, the increased size of the *modified Bai procedure,* unrestricted and restricted, is due to the multiple application of the test.

We conclude that, using these new methodologies and bootstrap critical values, the modified Bai test has reasonable power properties and a quite satisfactory performance in determining the number of breaks for the sample sizes, break sizes and degrees of persistence of interest in the long-run macroeconomics and international finance literature. In addition we show that restricted tests considerably increase power over unrestricted tests when the true alternative is restricted structural change and also preserve good size when the data includes one or two breaks of the same sign.

We next illustrate the practical importance of these results by using long-run real exchange rates. Purchasing Power Parity (PPP) is the hypothesis that real exchange rates exhibit reversion to a long-run constant mean while trend purchasing power parity (TPPP) postulates reversion to a long-run constant linear trend.

We start by documenting contradictory evidence of PPP from the application of unit root tests and multiple structural change tests. We argue that the evidence of long-run PPP or TPPP requires the rejection of the unit root null against a stationary or trend stationary alternative. On the other hand, while testing for structural change in the mean or trend of the real exchange rates, if we find a one-time change, or several changes occurring either in the same direction or opposite direction but of different magnitudes, long-run (T)PPP does not hold. If the changes are offsetting, the series returns to a constant mean (trend) and long-run (T)PPP holds.

The data covers real exchange rates for 16 industrialized countries with the US as the base country, starting from 1870 and ending in 1998. Using conventional (ADF) unit root tests, Papell and Prodan (2003) found evidence of PPP or TPPP by rejecting the unit root null hypothesis in favor of stationarity or trend-stationarity for 9 out of 16 countries. With unit root tests that both allow for structural change and maintain a long run mean or trend, they find evidence of a restricted PPP (or restricted TPPP) for 5 more countries. Therefore, using tests for unit roots, evidence for some variant of PPP can be found for 14 out of 16 countries.

Next, we test for possible mean shifts by applying methods proposed by Bai and Perron. Since the unit root tests allow for a maximum of two breaks, we impose the same constraints on the structural change tests. For 6 countries we find contradictions between unit root and structural change tests. We show that these contradictions are due to both the size distortions of the *Bai and Perron procedure* and to the low power of these tests on breaks of opposite sign. Using better sized tests that still maintain reasonable power (the *modified Bai procedure*) we are able to resolve only one contradiction. Using the *restricted procedure,* which further increases power on data that includes breaks of opposite sign while, in this particular case, maintaining a correct size, we are able to reconcile the results from unit root and structural change tests for 5 out of 6 countries.

The rest of the paper is structured as follows: Section 2 presents the Bai and Perron structural change model and documents the size distortions. In Section 3 we implement the Bai (1999) test for multiple structural changes and develop new methodologies which produce the best size and power properties of this test. We also present the results of simulations analyzing the size and the power of the test. Section 4 presents an empirical application, where we show the inconsistencies between unit root and structural change tests when testing for Purchasing Power Parity and we also demonstrate the properties of the new tests. Some concluding remarks are contained in Section 5.

2. Bai and Perron structural change tests

2.1. The Basic Model and Methodology

Bai and Perron (1998) consider multiple structural changes in a linear regression model, which is estimated by minimizing the sum of squared residuals.

We consider the following multiple linear regression with m breaks (m + 1 regimes):

$$y_t = x_t'\beta + z_t'\delta_j + u_t \qquad t = T_{j-1}+1,\ldots,T_j . \tag{1}$$

for $j = 1,\ldots,m+1$, $T_0 = 0$ and $T_{m+1} = T$.

In this model, y_t is the observed dependent variable at time t; x_t (p x 1) and z_t (q x 1) are vectors of covariates and β and δ_j (j = 1,…, m+1) are the corresponding vectors of coefficients; u_t is the disturbance at time t. In our study we consider models of both pure structural change, where all the regression coefficients are subject to change (p = 0), and partial structural change models, where only some of the coefficients are subject to change (parameter vector β is not subject to shifts and is estimated using the entire sample). Their models allow heterogeneity in the regression errors but they do not provide methods for estimating this heterogeneity.

Using Bai and Perron's method, based on the least-squares principle, we are able to estimate the regression coefficients along with the break points, when T observations are available. Bai and Perron (2003a) provide a detailed discussion. In the case of a pure structural break model (p = 0), where all coefficients are subject to change, for each possible $m-$partition $(T_1,\ldots,T_m)$ the least squares estimators of δ_j are obtained by minimizing the sum of square residuals. Then the estimated break points are the ones for which

$$(\hat{T}_1,\ldots,\hat{T}_m) = \arg\min_{T_{1,\ldots,}T_m} S_T(T_{1,\ldots,}T_m), \tag{2}$$

where $S_T(T_{1,\ldots,}T_m)$ denotes the sum of squared residuals. Since the minimization takes place over all possible partitions, the break-point estimators are global minimizers. Bai and Perron (2003a) use a very efficient algorithm for estimating the break points which is based on dynamic programming techniques. In the partial structural break model case (p > 0) the dynamic programming method to obtain global minimizers of the sum of squared residuals cannot be applied directly. They estimate the δ_js over the sub-samples defined by the break points, but the estimate of β depends on the final optimal partition $(T_{1,...,}T_m)$.[5]

A central result derived by Bai and Perron (1998) is that the break fraction $\hat{\lambda}_i = \hat{T}_i / T$ converges to its true value λ_i^0 at the fast rate T, making the estimated break fraction super-consistent. Therefore we can estimate the rest of the parameters, which converge to their true values at rate $T^{1/2}$, taking the break dates as known.

The Bai and Perron procedure allows for the estimation of the parameters and the confidence intervals under very general conditions regarding the structure of the data and the errors across segments. Their assumptions concerning the nature of the errors in relation to the regressors $\{x_t, z_t\}$ are of two kinds: First, when no lagged dependent variable is allowed in $\{x_t, z_t\}$, the conditions on the residuals are quite general and allow substantial correlation and heteroskedasticity.[6] The second case allows lagged dependent variables as regressors but no serial correlation is permitted in the errors $\{u_t\}$. In both cases, the assumptions are general enough to allow different distributions for both the regressors and the errors in each segment.[7]

The determination of the existence of structural change and the selection of the number of breaks depends on the values of various tests statistics when the break dates are estimated. Bai and Perron discuss three types of tests: a test of no break versus a fixed number of breaks, a double maximum test and a sequential test.

First, they consider a supF type test of no structural break (m = 0) versus m = k breaks. The test is $\sup F_t(k;q) = F_T(\hat{\lambda}_1,...,\hat{\lambda}_k;q)$ where $\hat{\lambda}_1,...,\hat{\lambda}_k$ minimize the global sum of squared residuals (according to (2)). Next, the null hypothesis of no structural break against an unknown number of breaks given some upper bound is tested by double maximum tests. Bai and Perron consider two statistics:

[5] In order to make this estimation possible, Bai and Perron modify a recursive procedure discussed in Sargan (1964). Briefly, they first minimize the sum of square residuals with respect to the vector of the changing parameters keeping β fixed and then minimize with respect to β keeping the vector of changing parameters fixed, and iterate. For appropriate initial values of β to start the iteration, convergence to the global minimum is attained. Bai and Perron (2003a) provide complete details of this estimation technique.

[6] In this case, an estimate of the covariance matrix robust to heteroskedasticity and autocorrelation can be constructed using Andrews's (1991) method, applied to the vector $\{z_t \hat{u}_t\}$. Bai and Peron use Andrews's (1991) data dependent method with the Quadratic Spectral kernel and an AR (1) approximation to select the optimal bandwidth. They also allow for pre-whitening as suggested in Andrews and Monahan (1992).

[7] They also consider cases where some constraints can be imposed on this general framework related to the distribution of the errors and the regressors across segments. A detailed description of different cases, when adding constraints, can be found in Bai and Perron (2002, 2003a). All these cases are allowed as options in the accompanying Gauss program.

$UD_{\max} = \max_{1 \le k \le M} \sup F_t(k;q)$, where M is an upper bound on the number of possible breaks, and also a version of this statistic, $WD_{\max}$, that applies weights to $\sup F_t(k;q)$ such that marginal p-values are equal across values of m. Finally, they proposed a test for ℓ versus $\ell+1$ breaks, labeled $F_t(\ell+1 \mid \ell)$. For this test the first ℓ breaks are estimated and taken as given. The statistic $\sup F_t(\ell+1 \mid \ell)$ is then the maximum of the F-statistics for testing no further structural change in the intercept against the alternative of one additional change in the intercept when the break date is varied over all possible dates.[8] The procedure for estimating the number of breaks suggested by Bai and Perron is based on the sequential application of the $\sup F_t(\ell+1 \mid \ell)$ test using the sequential estimates of the breaks.[9] The procedure can be summarized as follows: begin with a test of no-break versus a single break. If the hypothesis is rejected, one proceeds to test the null of a single break versus two breaks and so forth. This process is repeated until the statistics fail to reject the null hypothesis of no additional breaks. The estimated number of breaks is equal to the number of rejections.

Bai and Perron argue that even if the sequential procedure works best in selecting the number of breaks, there are certain configurations of changes (for example when two changes are present and the value of the coefficient returns to its original value after the second break) when this method breaks down.[10] Consequently, they propose *the Bai and Perron procedure,* which leads to the best results in empirical applications: First look at the sequential method, if 0 against 1 break is rejected, continue with the sequential method until the first failure to reject. If 0 against 1 is not rejected then look at $UD_{\max}$ or $WD_{\max}$ tests to see if at least one break is present. If the null of no breaks is rejected, then the number of breaks can be determined by looking at the sequential $\sup F_t(\ell+1 \mid \ell)$ statistics constructed using global minimizers for the break dates (ignore the test F(1|0) and select m such that the tests $\sup F_t(\ell+1 \mid \ell)$ are insignificant for $\ell \ge m$). All of the above mentioned test statistics have non-standard asymptotic distributions and Bai and Perron (2003b) provide the critical values for a trimming (ε) ranging from 0.05 to 0.25 and a value for q ranging from 1 to 10.

---

[8] The estimates $\hat{T}_i$ do not have to be the global minimizers of the sum of squared residuals, one can also use sequential one at a time estimates which allows the construction of a sequential procedure to select the number of breaks.

[9] They also discuss the use of two other common procedures to estimate the number of the breaks based on some information criterion: Bayesian Information Criterion (BIC) and modified Schwartz criterion (LWZ), which do not perform well in the presence of the serial correlation.

[10] In these cases, it is often difficult to reject the null hypothesis of 0 versus 1 break but it is not difficult to reject the null hypothesis of 0 versus a higher number of breaks.

2.2. Evidence of size distortions

There has been very little previous research that studies the size of tests for multiple structural changes. Bai and Perron (2001) provide size results for a range of data generating processes and for the various tests for structural change suggested. However, their results are limited to series with low persistence, specifically an AR(1) model with nuisance parameter $\alpha = 0.5$.

In this paper, we make the case that many relevant models from an economic point of view are highly persistent under the null of no structural change. Therefore we have a reason to doubt the accuracy of the critical values for such highly persistent processes. We illustrate this by extending the previous simulation evidence for the multiple structural changes tests to processes with larger roots. We also consider the size of the sequential method and the *Bai and Perron procedure* rather than focusing on the size of all individual tests.

2.2.1. Construction of the size simulations

We conduct Monte Carlo experiments for an AR (1) data generating process, with and without trend:

$$y_t = \alpha y_{t-1} + (\beta t) + \varepsilon_t \qquad (3)$$

We analyze the size of the test, specifically how well the procedure that selects the number of break points actually selects none when the data generating processes exhibit no structural change. Then the relationship between empirical and nominal test size is explored, varying the level of persistence measured by the autoregressive coefficient α =0, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99. In all of these cases the sample size is T=125 and 5000 replications are used with $\varepsilon_t = iidN(0,1)$. We choose T=125 because it provides a good approximation of the sample of long-run real GDP and real exchange rates data for industrialized countries.

We estimate two forms of equation (1):

$$\text{(a)} \quad y_t = c_j + \rho y_{t-1} + u_t \qquad (4)$$

$$\text{(b)} \quad y_t = c_j + u_t \qquad (5)$$

where $t = T_{j-1} + 1,...,T_j$, for $j = 1,...,m+1$, $T_0 = 0$ and $T_{m+1} = T$.

The first case (a) allows a lagged dependent variable as a regressor but no serial correlation is permitted in the errors $\{u_t\}$.This is a partial structural model where the breaks are assumed to be in the constant of the regression, while the autoregressive parameter, ρ, is estimated over the full sample, based on the optimal partition. The second case (b) is the case of a pure structural change model, when no lagged dependent variable is allowed (ρ = 0), the conditions on the residuals are general and allow substantial correlation.

We also consider two other similar cases, when estimating a trend-stationary AR(1) :

$$\text{(c)}\ \ y_t = c_j + \beta t + \rho y_{t-1} + u_t \qquad (6)$$

$$\text{(d)}\ \ y_t = c_j + \beta t + u_t \qquad (7)$$

The above cases are both partial structural change models: the breaks are assumed to be in the constant of the regression, but in (c) both, the trend and the autoregressive parameters (β and ρ) are estimated over the full sample and no serial correlation is permitted in the errors while in (d) only β is estimated over the full sample (ρ = 0), allowing for serial correlation in the errors.

Various versions of the tests can be obtained depending on the assumptions made with respect to the distribution of the data and the errors across segments. We will follow Bai and Perron (2003a) specifications. Therefore in our size simulations, due to the highly persistent generated data, we correct for serial correlation (for simplicity we denote it cor_u = 1) in cases where we do not allow for a lagged dependent variable. We consider cases where we allow for heterogeneous variances of the residuals (het_u = 1) and also cases where we restrict the residuals to have the same variance throughout (het_u = 0) [11]. We allow for different distributions of the data across segments in all cases.[12]

The size of the multiple structural change tests is also sensitive to the choice of the trimming parameter, ε . Following Bai and Perron's (2003b) recommendation to achieve tests with correct size in finite samples, when allowing for heterogeneity across segments and errors of the estimated regression model, we use a value of the trimming ε = 0.15 with M = 5 and ε = 0.20 with M = 3 (maximum number of breaks). Even though the size of the test is improved in some cases with ε = 0.20, there is a trade-off when using a higher value than 0.15 for trimming; not only do we lose important information from the data but the maximum number of breaks allowed is smaller.

2.2.2. Simulation results

Results for the model without trend are shown in Table 1 and 2 and those for the model with trend in Table 3 and 4. We present the rejection rate under the null hypothesis of no structural change for

---

[11] Bai and Perron (2001) show that correcting for heterogeneity and serial correlation improves the power of the tests and the accuracy in the selection of the number of the breaks.

[12] Allowing for a common distribution of the errors across segments leads to tests with worse properties, even if the data indeed has an invariant distribution.

a range of α from 0 to 0.99. The *i.i.d*. errors data is a base case that is compared with cases where the data is persistent.

We report size results for tests with nominal size of 5%. The Bai and Perron recommendation is to correct for heterogeneity in the variance of the residuals across segments in order to improve the power of the tests and the accuracy in the selection of the number of breaks. We therefore focus on cases where het_u = 1 (we allow for heterogeneity of the residuals) and ε = 0.15[13]. We report both the size of the sequential method and the size of the *Bai and Perron procedure*. In each table (1 to 4) we report size results for the sequential method (total size) and for the *Bai and Perron procedure* (the size corresponding to each number of breaks as well as the total size). The second column, in which the size result for one break is reported, is common to both methods.[14]

We begin by considering the experiment where we choose the number of breaks with the sequential method. First we consider the case with non-trending stationary data. In Table 1 we estimate a model where we account for serial correlation in a direct parametric fashion (we introduce a lagged dependent variable). For α = 0.9 and het_u = 1 the rejection rate for the nominal 5% test based on the Bai and Perron asymptotic critical values is 21.88%. Similarly, if het_u = 0 the rejection rate test is 20.11%. Therefore, in highly persistent series, restricting het_u to be 0 does not seem to improve the size. In Table 2 we estimate a model where we use an indirect non-parametric correction to take in account these dynamic effects. For almost all levels of persistence the size distortions exceeds the values obtained in the previous case. For α = 0.9 the rejection rate is 27.88% if het_u = 1 and 21.42% if het_u = 0. This shows, in line with previous literature, the tendency of heteroskedasticity and autocorrelation consistent (HAC) robust tests to over reject, sometimes substantially, under the null hypothesis.

Bai and Perron (2003a) argued that, even if the asymptotic theory is valid only for cases with non-trending data, one can safely use the same critical values for the case with trending data. We therefore consider cases where the data generating process includes a time trend, which we estimate over the full sample. In cases where we account for serial correlation in a parametric fashion (Table 3), we have extreme size distortions. For instance, for α = 0.9 and het_u = 1, the rejection rate for the nominal 5% test is 42.54%. Restricting the heterogeneity of the residuals to be the same (het_u = 0) does not seem to improve the size: for α = 0.9, the rejection rate is 41.70%. Surprisingly, if we account for serial correlation in a non-parametric way (Table 4), the size distortions become smaller.[15] For α = 0.9, the rejection rate is 22.26%.

---

[13] The results for ε = 0.20 and het_u=0 are also reported in the tables.

[14] In order to produce the best size when using the non-parametric method and provide comparable results with Bai and Perron (2001) we pre-whiten the residuals as suggested by Andrews and Monahan (1992).

[15] Following Bai and Perron (2003a) specifications for partial structural change models when allowing for serial correlation, we only consider the case where het_u = 0.

We next consider the experiment where we choose the number of breaks with the *Bai and Perron procedure*. We first generate non-trending data. In Table 1 we estimate a model where we take in account for serial correlation in a direct parametric fashion. For α = 0.9 and het_u = 1 the rejection rate for the nominal 5% test is 37.30% for het_u = 1 and 35.2% for het_u = 0. In Table 2, where we estimate a model that uses a non-parametric correction, for α = 0.9 the rejection rate is 48.74% if het_u = 1 and 43.34% if het_u = 0. Second, we consider cases where the data generating process includes a time trend. In cases where we account for serial correlation in a parametric fashion (Table 3), for α = 0.9, the rejection rate for the nominal 5% test is 59.28% for het_u = 1 and 59.04% for het_u = 0. Accounting for serial correlation in a non-parametric way (Table 4), for α = 0.9, the rejection rate is 32.70%. It is evident that the *Bai and Perron procedure* severely increases the size distortions over the sequential method, the magnitude depending on the model and the persistence level.

The size distortions increase as the level of persistence rises and are very large for coefficients near unity. For instance, for α = 0.99 and het_u = 1, the size of the sequential method ranges between 47.08% and 48.12% when testing non-trending data and between 26.42% and 61.46% when testing trending data. For the same parameter, α = 0.99, the size of the *Bai and Perron procedure* ranges between 64.12% and 68.66% when testing non-trending data and between 37.66% and 73.78% when testing trending data. If we impose the same variance of the residuals across segments (het_u = 0) we obtain in all cases a slightly better size.[16] If a trimming (ε) of 0.20 is used the size distortions are slightly smaller, but this is partially due to the fact that we estimate 3 versus 5 breaks.

We conclude that, for each model, the sequential method has a strong tendency to spuriously reject the null hypothesis of structural change for realistic values of α and the sample size. Moreover, while improving the power on breaks of opposite sign, the *Bai and Perron procedure* severely increases the evidence of size distortions compared with the already oversized sequential method.

3. Bai multiple structural change test

3.1. The model

3.1.1 The unrestricted version

In the previous section we showed that the sequential method, as well as the *Bai and Perron procedure,* suffers from severe size distortions when testing highly serial correlated data. We intend to

---

[16] On the other hand, the Bai and Perron recommendation is to correct for heterogeneity in the variance of the residuals across segments to improve the power of the tests and the accuracy in the selection of the number of breaks.

correct this problem by obtaining the best possible size for multiple structural change tests while maintaining good power properties of these tests on data that includes any type of change. Consequently, we implement a similar method, the Bai (1999) likelihood ratio type test. Its essential feature lies in its easy applicability and the use of straightforward methods in order to calculate bootstrap critical values. It allows for lagged dependent variables and it is well suited to test multiple changes in polynomial trends, both in the intercept and slope of the trend function.[17]

The estimated model is similar to Bai and Perron, equation (1). It also allows for pure and partial structural changes. For the purpose of this study we estimate the following specification of equation (1)[18]:

$$y_t = c + (\beta t) + \sum_{l=1}^{m} \theta_\ell DU_{\ell t} + \sum_{i=1}^{k} \rho_i y_{t-i} + u_t \tag{8}$$

The lag length, k, is chosen by Schwarz Information Criterion (SIC), which involves minimizing the function of the residual sum of squares combined with a penalty for a large number of parameters in the previous regressions. We set the upper bound on the number of structural changes (m) to 5. We consider two cases: in the first case β =0 and the autoregressive parameter (ρ) is estimated over the full sample while in the second case both, the trend and the autoregressive parameters (β and ρ), are estimated over the full sample.

While the estimated model is similar to Bai and Perron, the testing procedure is different. Bai and Peron propose a sequential test of ℓ versus $\ell+1$ breaks, but they estimate any additional break point conditional on the previously obtained break points, splitting of the sample. The disadvantage of this method is that one quickly runs out of degrees of freedom, which is particularly important for structural change models. Bai's testing methodology has a likelihood ratio interpretation: the model is estimated optimally under both null and the alternative hypothesis, which means that ℓ breaks under the null and $\ell+1$ breaks under the alternative are estimated simultaneously. When $\ell = 0$, the test reduces to the usual test of no change against a single change. When the test is performed repeatedly while augmenting the value of ℓ, the number of break points can be consistently estimated. This test has a different limiting distribution than the Bai and Perron's sequential test and it is conceptually much simpler to apply.

Because the Bai test has a sequential interpretation, it suffers from the same power problems as the Bai and Perron sequential method. In certain configurations of changes, particularly breaks of opposite

[17] The limiting distribution is characterized and analytical expressions for critical values are derived in the case of trending data. Bai and Perron (2003a) claim that simple modifications can be applied to their method to deal with trending regressors. The asymptotic theory of their tests is valid only for the case of non-trending data, but it is fairly similar to the trending data case. Hence, they argue that one can use the same values in both cases: non-trending and trending data.

[18] This study considers only changes in the intercept. The behavior of these tests when the process includes changes in the slope of the trend function is subject of future research.

sign, the Bai sequential method is unable to reject the null hypothesis of 0 versus 1 break but it is not difficult to reject the null hypothesis of 0 versus of a higher number of breaks. In order to improve the power, while minimizing size distortions, we extend the Bai test in several directions. Similar to the *Bai and Perron procedure,* we propose the following methodology: First, look at the sequential method. If 0 against 1 break is rejected, continue with the sequential method until the first failure to reject. If 0 against 1 is not rejected then test the hypothesis of no break versus a fixed number of breaks. If any $\sup F_t(\ell)$ (for $\ell = 0$ versus $\ell = 1, \ldots,$ k breaks, where k is maximum number the breaks considered) is significant, then the number of breaks can be decided upon a sequential examination of the $F_t(\ell+1|\ell)$ for $\ell = 1, \ldots,$ k breaks. Select the number of breaks, m, such as the tests $F_t(\ell+1|\ell)$ are insignificant for $\ell \geq m$.[19] We will call it the *modified Bai procedure.*

3.1.1 The restricted version

The rationale for moving beyond the sequential procedure, in both the Bai and Perron and the Bai test, is to improve power when the data contains breaks of opposite sign. We now propose a restricted version of Bai test that has the potential to produce further power improvements. If under the alternative the series includes breaks of opposite sign, designing a specific restriction which captures this configuration raises the possibility of improving power on that specific data. Consequently, we restrict the coefficients on the dummy variables that depict the breaks in equation (8) to produce a constant mean or trend:

$$\sum_{\ell=1}^{m} \theta_\ell = 0 \tag{9}$$

This imposes the restriction that the mean following the last break is equal to the mean prior the first break. We propose the following methodology in order to choose the number of restricted breaks: test the hypothesis of no break versus a fixed number of restricted breaks. If any $\sup F_t(\ell)$ (for $\ell = 0$ versus $\ell = 2, \ldots,$ k breaks, where k is maximum number the breaks considered) is significant, then the number of breaks can be decided upon a sequential examination of the $F_t(\ell+1|\ell)$ for $\ell = 2, \ldots,$ k breaks and $F_t(2|0)$. Select the number of breaks, m, such as the tests $F_t(\ell+1|\ell)$ and $F_t(2|0)$ are insignificant for $\ell \geq m$. We will call it the *restricted procedure*.

[19] We also considered other variations, but later simulations showed that this methodology maximizes size and power performance.

3.2. Finite sample performance: Size and Power analysis

3.2.1. Size correction

As in the previous case, we consider an AR (1) data generating process, with and without trend (equation 3), varying the level of persistence measured by the autoregressive coefficient α = 0, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99. In all of these cases the sample size is T=125 and 5000 replications are used with $\varepsilon_t = iidN(0,1)$. We consider estimating four specifications: First, we estimate equation (8), with and without trend. Next we estimate the same equations, but we also add the restriction in equation 9.For all these cases we use a trimming (ε) of 0.05.

Critical values are calculated using parametric bootstrap methods.[20] Generally we assume that the underlying process follows a stationary finite-order autoregression of the form:

$$A(L)y_t = e_t, \tag{10}$$

$e_t \sim iidN(0,1)$ with $E(e_t) = 0$ and $E(e^2{}_t) < \infty$. $Y = (y_{1,...,}y_T)'$ denotes the observed data.[21] $A(L)$ is an invertible polynomial in the lag operator. The AR(p) model may be bootstrapped as follows: First determine the optimal AR(p) model, using the Schwartz criterion. Next, estimate the parameters $\hat{A}(L)$ for the optimal model. Following, to determine the finite-sample distribution of the statistics under the null hypothesis of no structural change, use the optimal AR model with $iidN(0,\sigma^2)$ innovations to construct a pseudo sample of size equal to the actual size of the data, where σ^2 is the estimated innovation variance of the AR model. Then calculate the bootstrap parameter estimates: $\hat{A}*(L)$ and compute the statistics of interest. The critical values are taken from the sorted vector of 5000 replicated statistics.

We present the results in Table 5. If we use the sequential method and bootstrap critical values, we obtain a correct size of 5%. The use of our newly developed *modified Bai procedure*, however, leads to an increased size, which is due to the multiple application of the test. In both, the unrestricted and restricted versions of the tests, the size associated with the use of bootstrap critical values is approximately 10%, regardless of the value of the nuisance parameter α.[22] In the case of the unrestricted

[20] Berkowitz and Killian (2001) provide a detailed explanation of the parametric bootstrap procedure. They also claim that for many applications in time series econometrics parametric methods are preferable to non-parametric methods.

[21] Further research is needed to assess the sensitivity of our results with respect to departures from Normality.

[22] It is possible to choose critical values that are smaller than 5%, for each application of the test, in order to obtain an actual size of 5% of the overall procedure, unrestricted or restricted. This, however, will result in a loss of power, which we intend to investigate in future research. For the purpose of this study we limit ourselves to bootstrap critical values of 5%, which assures a correct size when testing for two restricted breaks in the empirical application.

model and non-trending data, for α =0.9, the rejection rate is 11.5%. For trending data and a persistence of 0.9, the actual size is 10.3%. We find that the restricted tests perform better: for non-trending data and α =0.9, the rejection rate with nominal size of 5% is 9.8%. The restricted tests perform even better in the case of trending data, with an actual size of 8.9%.

These results demonstrate that the bootstrap distribution can be a much better approximation to the finite-sample distribution than the asymptotic distribution when the persistence is high. In contrast with the *Bai and Perron procedure* the size does not increase as the persistence of the data generating process rises, even as the autoregressive coefficient becomes close to unity. [23]

3.2.2. Size adjusted power analysis

We have shown that the size distortions of multiple structural change tests can be corrected by the use of appropriately adjusted critical values. The issue becomes the loss in power implied by such gains in the size performance of the test. We proposed two methods to improve the power of the Bai test. First, we develop a new procedure which improves the power of the sequential method, which we call the *modified Bai procedure*. Second, we develop a restricted version of this new procedure which further improves the power when the breaks have certain configurations, as breaks of opposite sign (*the restricted procedure*). Bai (1999) reports power simulation results for his sup LR test in comparison with Bai and Perron's conditional test $\sup F_t(\ell+1\,|\,\ell)$, using asymptotic critical values. He points out that even if the conditional procedure does a satisfactory job, the sup LR tests show an improved performance. His experiment analyzes the sequential procedure and it is restricted to a data generating process with a persistence of only 0.5. In contrast, we conduct power experiments on highly persistent data using our newly developed procedures.

Our power simulation experiments address the following issues: (a) comparison between the *modified Bai procedure* and the *restricted procedure* performances and (b) size adjusted power as the function of the level of persistence. [24]

Within Monte Carlo experiments we consider the following data generating processes (DGP):

$$y_t = \alpha y_{t-1} + (\beta t) + \theta DU_{it} + \varepsilon_t \, , \qquad (11)$$

The power of the multiple structural change tests is investigated by constructing experiments with artificial data under a true alternative hypothesis where the data process is stationary (or trend-stationary), allowing for one, two or four changes in the intercept (i =1, 2 and respectively 4 in equation 11).

[23] Inoue and Kilian (2002) investigate bootstrapping autoregressive processes with possible unit roots.
[24] Since the sequential method has no power on data that includes breaks of opposite sign, we do not report simulation results for the original sequential Bai method.

The power of a test is normally analyzed by tabulating how often the null is rejected when it is false. When testing for multiple structural change tests the definition of the power is ambiguous. The power can be defined as (1) finding at least one break regardless of the number of breaks included in the DGP, (2) finding at least as many breaks as are included in the DGP or, (3) finding the exact number of breaks included in the DGP. We analyze the power of the test by tabulating how often the procedure selects the exact number of break points present in the DGP.

The generated data is based on economically plausible assumptions. We specify the nuisance parameter α = 0.6, 0.7 and 0.8, the trend slope $\beta = 0.01$ and the magnitude of the breaks as being 0.3, corresponding to a standard deviation of the residuals of 0.223, which covers most of the cases found in the annual real exchange rate data analyzed in the next section.[25] The timing of the break is set at the 1/3 and 2/3 of the sample. In all of these cases the sample size is T=125 and 5000 replications are used with $\varepsilon_t = iidN(0,0.223)$. We calculate bootstrap critical values as shown in section *3.2.1* and report results for the 5% nominal size. Two versions of the test are estimated: the unrestricted (equation 8) and the restricted model (equations 8 and 9).

To address these issues we consider a stationary DGP where we account for different configuration of breaks in the intercept (equation 11). There are some general features of the *modified Bai procedure* and the *restricted procedure*: As the errors become less persistent there is a monotonic increase in power. Also, the power of the test decreases as the number of the breaks included in the data generating process increases. Due to the lack of space, we mainly report power results for α = 0.7 (detailed results for different values of α are presented in each table).

The *modified Bai procedure* (see Table 6) has very good power when the DGP is stationary with one break (93.5%) and moderate power with two breaks of the same sign (59.1%). Also the power of the *modified Bai procedure* is fairly good (72.1%) when data is generated with two breaks of opposite sign and lower (37.4% and 42.2% for case 1 and 2, respectively) when data is generated with four opposite breaks.

The *restricted procedure* has very low power when the generated data includes one permanent change or two changes that occur in the same direction. This is important for empirical models because the tests will not provide evidence of restricted structural change when, in fact, the actual structural change is not consistent with the restriction. If we consider a process that includes breaks that are equal and of opposite sign, we find an increase in the power of the *restricted procedure* over the *modified Bai procedure* of more than 10%, especially for cases with highly persistent errors.

[25] We therefore assume a change in the mean of approximately 1.3 standard deviations of the residuals. Perron and Vogelsang (1992) report a change in the mean of 1.2 standard deviations, arguing that this value is likely in practical instances.

We proceed to apply the previous tests, including a time trend in the estimation, to trend stationary data that also include one, two or four changes in the intercept (Table 7). The power of the *modified Bai procedure* and the *restricted procedure* is smaller than in case of non-trending generated data. Considering $\alpha = 0.7$, the *modified Bai procedure* has moderate power when the DGP is stationary with one break (53.8%) and low power with two breaks of the same sign (30.3%). Its power is even lower when data is generated with two breaks of opposite sign (20.5%). When the data includes four breaks of opposite sign the power of the test is insignificant, only 14.5% and 12.2% for case 1 and 2 respectively. In this case, the procedure mistakenly finds 2 breaks with a higher probability than finding the correct number of breaks (25.1% and 30.9% for case 1 and 2, respectively).

The *restricted procedure* substantially improves power over the *modified Bai procedure* if the process is consistent with breaks that are equal and of opposite sign. With two breaks, the power of the *restricted procedure* increases by approximately 50% in comparison with the *modified Bai procedure* for all levels of persistence (for $\alpha = 0.7$ the power of restricted tests is 70.6%). With four breaks, the power of the *restricted procedure* increases by approximately 30% over the *modified Bai procedure*. The *restricted procedure* has insignificant power when the generated data includes two changes that occur in the same direction, but it has small power when data includes one break.

4. Empirical application: Purchasing Power Parity

A large literature has emerged on testing the long-run validity of purchasing power parity (PPP), or equivalently the stationarity of real exchange rates, using modern time-series econometrics techniques. Inspired by the obvious failure of PPP to hold in the short run following the end of the Bretton-Woods system, testing for long-run PPP became synonymous with testing the unit root null against the stationary alternative in real exchange rates (nominal exchange rates adjusted for national price differentials). The post Bretton-Woods period, however, is too short for univariate methods to be informative. Attention has moved to panel methods for monthly or quarterly short-horizon and univariate methods for annual long-horizon data.

While the investigation of time series properties of long-horizon real exchange rates has gained an important place in PPP studies, the analysis of a century (or more) of real exchange rates raises both conceptual and methodological issues. Following Cassel, the dominant view of PPP is that of long-run mean reversion. From that perspective, evidence of PPP can be found by rejecting the unit root null in real exchange rates against a level stationary alternative. With long-horizon data, however, it becomes necessary to consider theories of long-run real exchange rate movements. The best known of these theories, the Balassa-Samuelson model, is usually interpreted as implying trending real exchange rates.

With this interpretation, evidence of the Balassa-Samuelson view can be found by rejecting the unit root null in real exchange rates against a trend stationary alternative. We call this Trend Purchasing Power Parity.

In this paper we focus on the investigation of the time series properties of long-horizon real exchange rates. Many papers that test the unit root null against level and/or trend stationary alternative for real exchange rates have been recently published. 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. Cuddington and Liang (2000) and Lothian and Taylor (2000) investigate the implications of allowing for a linear trend in the Lothian and Taylor (1996) data. 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.

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.[26] The real dollar exchange rate is calculated as follows:

$$q = e + p^* - p, \tag{11}$$

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 test the validity of long-run purchasing power parity, using two widely-used methods: unit root and structural change tests. The evidence of PPP or TPPP requires the rejection of the unit root null against a stationary or trend stationary alternative. When testing for structural change, however, if we find a one-time change in the mean or trend of the real exchange rate, long-run (T)PPP does not hold. The situation becomes more complicated when real exchange rates experience multiple structural changes in the mean or trend. If the changes are offsetting, the series returns to a constant mean (trend) and long-run (T)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

[26]The countries are Australia, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom and United States.

(trend) and long-run (T)PPP does not hold. The results from these two tests, unit root and structural change, have to lead to the same conclusion in order to provide a unified evidence of PPP or TPPP.

Using ADF tests, Papell and Prodan (2003) find evidence of PPP for 8 out of 16 countries at the 5% level, Belgium, Germany, Finland, France, Italy, Norway, Spain, and Sweden, by rejecting the unit root null in favor of stationarity. Incorporating time trends, consistent with the Balassa-Samuelson theory, evidence of TPPP at the 5% level is found for one additional country, Australia.[27] The use of more powerful DF-GLS tests (with and without a time trend) and MAIC lag selection does not increase the number of rejections[28].

In order to test for PPP (or TPPP) while allowing for structural change, Papell and Prodan (2003) 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. Using restricted unit root tests, 5 more rejections are added to the previous results: evidence of PPP restricted structural change for Portugal and United Kingdom and evidence of TPPP restricted structural change for Denmark, Japan and Switzerland. Power simulations show that the restricted tests have no power when the process is inconsistent with the PPP or TPPP hypotheses, therefore these rejections represent strong evidence of PPP or TPPP. [29]

Using conventional and restricted unit root tests, evidence of some variant of PPP can be found for 14 of the 16 countries. The Netherlands experiences quasi purchasing power parity, reversion to a changing mean with one structural change, and Canada is the only country where no evidence of any variant of PPP was found. Results for ADF and restricted structural change tests are cited in Table 8. [30]

In this paper we investigate the (T)PPP hypothesis by using tests for multiple structural changes with long-run real exchange rate data. Since the unit root tests allow for a maximum of two breaks, we impose the same constraints on the structural change tests. While the *Bai and Perron procedure* still exhibits large size distortions and the *modified Bai procedure* remains somewhat oversized, the restricted procedure with two breaks has correct size.[31] We first use the *Bai and Perron procedure* and find six contradictions between the results of unit root and structural change tests[32]. Due to the severe size

[27] West (1987) shows that ADF tests that do not include a time trend have zero asymptotic power if the series is trend stationary. Therefore we can discriminate between evidence of PPP and TPPP.

[28] Using DF-GLS tests and MAIC lag selection, Lopez, Murray and Papell (2003) also found evidence of (T)PPP for 9 out of 16 countries. In comparison with Papell and Prodan (2003), they find evidence of PPP for Australia, Belgium, Germany, Finland, Italy, Netherlands, Spain, Sweden and UK and no evidence of TPPP. Due to the long-span of data there is no increase in power using DF-GLS test over conventional unit root tests.

[29] If the data is trending with breaks of equal and opposite sign (TPPP restricted structural change), tests that incorporate PPP restricted structural change have very low power. In accord with previous ADF tests, they can discriminate between PPP and TPPP restricted structural change.

[30] For a detailed discussion regarding quasi purchasing power parity see Hegwood and Papell (1998).

[31] The size of the *Bai and Perron* and the *modified Bai procedure* with two breaks is between the column marked "Sequential" and "Total" in Tables 1-5.

[32] We report results for the non-parametric procedure with a 5% level of significance and a trimming (ε) of 0.15.

distortions of the *Bai and Perron procedure*, we next use the *modified Bai procedure*. Finally, in order to improve the power of the *modified Bai procedure* on breaks of opposite sign, we apply the *restricted procedure.*[33]

We test the 15 cases where we previously found evidence of regime wise (trend) stationarity (Canada is the only country which we do not consider furthermore). The results of these tests are presented in Table 9, 10 and 11.

We divide the evidence found for the real exchange rates into 5 categories: PPP, TPPP, restricted PPP, restricted TPPP and quasi PPP. Within each category, we investigate the consistency of unit root and structural change tests results when testing for a variant of PPP.

We start with the eight countries where we found evidence of PPP with unit root tests. Using the *Bai and Perron procedure,* we find that only one country, France, experiences (one) structural change. This result implies that we find both evidence of PPP and structural change inconsistent with PPP. Using the *modified Bai procedure* we also find one break. Finally, using the *restricted procedure* we solve this contradiction, finding evidence of restricted structural change consistent with PPP (Figure 1A). For the country where TPPP holds, Australia, we do not find any evidence of structural change.

We next consider the two countries where we find evidence of PPP restricted structural change: Portugal and the United Kingdom. Using the *Bai and Perron procedure* we find contradictions in both cases: In the case of Portugal we find evidence of one break, inconsistent with the evidence of PPP restricted structural change. Using the *modified Bai procedure,* we find no evidence of breaks. Applying the *restricted procedure*, we find evidence of restricted structural change, therefore solving the contradiction. We find no evidence of structural change evidence for United Kingdom with the *Bai and Perron procedure*, which is not consistent with the results of the restricted unit root tests. Using the *modified Bai procedure,* we find evidence of two breaks of opposite sign. While this result is potentially consistent with restricted structural change, the breaks could have different magnitudes. We finally solve the contradiction, finding evidence of restricted structural change with the *restricted procedure* (Figure 1A).

We then consider the three countries where we find evidence of TPPP restricted structural change: Japan, Denmark and Switzerland. For all these countries, the *Bai and Perron procedure* leads to inconsistent results: For Japan and Denmark, we do not find evidence of structural change. In these cases, correcting for size with the *modified Bai procedure* does not affect the result. Increasing the power with the *restricted procedure* we solve the contradiction for Japan, finding evidence of restricted structural change, but we still find no evidence of structural change in the case of Denmark. Second, for

[33] For the *modified Bai procedure* and the *restricted procedure* we report results for a 5% level of significance and a trimming (ε) of 0.05.

Switzerland, we find evidence of two breaks of opposite sign. If the breaks of opposite sign are equal, the result would be consistent with the restricted unit root test result. Using the better sized test, the *modified Bai procedure,* we find evidence of one structural change, which leads to the conclusion that the previous finding of two breaks of opposite sign could be the result of the *Bai and Perron procedure's* size distortion. Furthermore, using the *restricted procedure* we find evidence of restricted structural change, solving the contradiction (Figure 1B).

For the case of Netherlands, where we found evidence of quasi purchasing power parity (with one structural change) using unit root tests, we also find evidence of one structural change using both the *Bai and Perron* and the *modified Bai procedure*. The *restricted procedure* does not provide any evidence of restricted structural change, which is again consistent with the quasi PPP hypothesis of one structural change.

Comparing the results from unit root and structural change tests, we find 6 contradictions among the 15 countries considered.[34]. Since the conventional and the restricted unit root tests have a good size on data that includes structural changes which are inconsistent with PPP hypothesis, we conclude that these contradictions are due either to the size distortions of the *Bai and Perron procedure* or to the low power of the *modified Bai procedure* on data with breaks of opposite sign. Using the *restricted procedure*, which increases the power on data that includes breaks of opposite sign while maintaining a correct size, we are able to solve 5 out of 6 contradictions.

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 both the *restricted procedure*. 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, France 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.

[34] In cases as Finland and Italy we find one spike (breaks of opposite sign which are very close to each other). We also find in both cases evidence of restricted structural change therefore we do not consider these results as a contradiction.

5. Conclusion

It is common in empirical work to test for one or more structural changes when analyzing macroeconomic and international finance data. Bai and Perron (1998) consider estimating and testing for multiple structural changes, proposing a sequential procedure which finds breaks by splitting of the sample. This test is widely used, partially due to the accompanying efficient algorithm that estimates multiple structural changes, based on the principle of dynamic programming. In later studies, they propose a procedure which improves the power of the sequential method on special configuration of changes, mostly breaks of equal and opposite sign (we call it the *Bai and Perron procedure*).

The purpose of this paper is (1) to identify a potential serious problem in the use of these tests and (2) to propose methods to solve this problem. We show that the Bai and Perron sequential tests suffer from severe size distortions, rejecting the no-structural-change null hypothesis too often when it is true, if the data is highly persistent. The size distortions are made much worse by the use of the *Bai and Perron procedure.* For both the sequential method and the *Bai and Perron procedure,* the size distortions become more severe as the persistence increases. Since many processes of interest in empirical macroeconomics and finance tend to be highly persistent, our results of size distortions are of immediate practical interest. If there is a possibility that the data contain breaks of equal and opposite sign, one is faced with a trade-off between tests with very low power to detect such breaks (the sequential test) and tests that are very badly oversized (the *Bai and Perron procedure*).

In order to correct for size distortions, we implement the Bai (1999) likelihood ratio test. Its essential feature lies in its easy applicability, allowing for straightforward calculation of bootstrapped critical values. We show how finite sample bootstrapping methods can achieve correct size for the sequential results regardless of the persistence in the data. However, similar to the Bai and Perron tests, the sequential method has low power to detect breaks of equal and opposite sign. We extend the Bai test in two directions: First, we propose a methodology, similar to the *Bai and Perron procedure*, which we call the *modified Bai procedure.* While the size of the sequential tests is 5%, the size of this procedure is approximately 10% - 11%, due to the multiple application of the test.

Second, in order to further improve the power on these specific changes, we propose a restriction that specifically models data including two or more changes which imply mean or trend reversion, which we call *restricted procedure.* The size of the restricted procedure is approximately 10%. The size of both the *modified Bai procedure* and the *restricted procedure* does not increase with the persistence of the data. The power simulations show a reasonable power performance for the *modified Bai procedure* and a substantially increased power for the *restricted procedure,* especially in the case of trending data.

We illustrate the practical importance of these results by analyzing long-run real exchange rates, using the longest span of available historical data for 16 industrialized countries with the US as the base country. We test the validity of long-run (trend) purchasing power parity, using two widely-used methods: unit root (conventional and restricted structural change) tests and structural change tests. Evidence of PPP or TPPP requires the rejection of the unit root null against a stationary or trend stationary alternative. In addition, when testing for structural change in the mean of real exchange rates, if we find a one-time change, or several changes occurring either in the same direction or opposite direction but of different magnitudes, long-run (T)PPP does not hold. If the changes are offsetting, the series returns to a constant mean (trend) and long-run (T)PPP holds.

Comparing results from the unit root tests and the Bai and Perron structural change tests, we find 6 contradictions among the 15 countries: France, Portugal, UK, Japan, Switzerland and Denmark. These contradictions are due to both size distortions and to the low power of the tests with breaks of equal and opposite sign. Using better sized tests (*modified Bai procedure*) we are able to resolve only 1 out of 6 contradictions. By using tests which increase the power on data that includes breaks of opposite sign while maintaining a correct size (*restricted procedure*), we are able to resolve 5 out of 6 contradictions.

References

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

Andrews, Donald W. K. (1991), “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59, 817-958.

Andrews, Donald W. K. and J. C. Monahan (1992), “An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator,” *Econometrica*, 60, 953-966.

Andrews, Donald W. K. (1993), “Tests for Parameter Instability and Structural Change with Unknown Change Point,” *Econometrica*, 61, 821-856.

Andrews, Donald W. K. and Werner Ploberger (1994), “Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative,” *Econometrica*, 62, 1383-1414.

Bai, Jushan and Pierre Perron (1998), “Estimating and Testing Linear Models with Multiple Structural Changes”, *Econometrica*, 66, 47-78.

Bai, Jushan and Pierre Perron (2001), “Multiple Structural Change Models: A Simulation Analysis,” forthcoming in *Econometric Essays*, D. Corbea, S. Durlauf and B.E. Hansen (eds.), Cambridge University Press.

Bai, Jushan and Pierre Perron (2003), “Computation and Analysis of Multiple Structural Change Models,” *Journal of Applied Econometrics,* 18, 1-22.

Bai, Jushan and Pierre Perron (2003), “Critical Values for Multiple Structural Change Tests,” *Econometrics Journal,* 6, 72-78.

Bai, Jushan (1999), “Likelihood Ratio Tests for Multiple Structural Changes,” *Journal of Econometrics*, 91, 299-323.

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

Berkowitz, Jeremy and Lutz Killian (2000), “Recent Developments in Bootstrapping Time Series,” *Econometric Reviews*, 19, 1-54.

Chow, Gregory C. (1960), ”Tests of Equality Between Sets of Coefficients in Two Linear Regressions,” *Econometrica*, 28, 591-605.

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

Diebold, Francis X. and C. Chen (1996), "Testing Structural Stability with Endogeneous Break Point: A Size Comparison of Analytic and Bootstrap Procedures," *Journal of Econometrics,* 70, 221-241.

Diebold, Francis X. and Abdelhak S. Senhadji (1996), “Deterministic vs. Stochastic Trend in U.S. GNP, Yet Again,” *American Economic Review*, 86, 1291-1298.

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

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

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

Inoue, Atsushi and Lutz Kilian (2002), "Bootstrapping Autoregressive Processes with Possible Unit Roots," *Econometrica*, 70, 377-391.

Lopez, Claude, Christian J. Murray and David H. Papell (2003), "State of the Art 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.

Papell, David H. and Ruxandra Prodan (2003), "Long Run Purchasing Power Parity: Cassel or Balassa-Samuelson?" working paper, University of Houston.

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

Rudebusch, Glenn D. (1993), "The Uncertain Unit Root in Real GNP," *The American Economic Review*, 83, 264-272.

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

Stock, James H. (1994), "Unit Roots, Structural Breaks and Trends," Handbook of Econometrics, Volume IV, R.F. Engle and D.L. McFadden (eds).

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

Vogelsang, Tim J. (1997), "Wald-Type Tests for Detecting Breaks in the Trend Function of a Dynamic Time Series," *Econometric Theory*, 13, 818-849.

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

Table 1. Size of the sequential method and the *Bai and Perron procedure*: the non-trending parametric case

Estimated model: $y_t = c_j + \rho y_{t-1} + u_t$

(a) cor_u = 0, het_u = 1

α	Sequential	Bai and Perron procedure					
	Total	1break	2 breaks	3 breaks	4 breaks	5 breaks	**Total**
				ε **= 0.15**			
0	**6.74**	6.54	1.86	0.12	0.00	0.00	**8.52**
0.5	**8.68**	8.26	2.78	0.32	0.10	0.00	**11.46**
0.6	**9.50**	8.90	3.54	0.56	0.14	0.00	**13.14**
0.7	**10.86**	10.28	4.52	0.92	0.28	0.02	**16.02**
0.8	**13.80**	12.74	7.12	1.64	0.60	0.02	**22.12**
0.9	**21.88**	19.56	13.04	3.56	1.06	0.08	**37.30**
0.95	**32.46**	27.32	17.84	5.40	1.50	0.12	**52.18**
0.99	**47.08**	36.98	20.62	4.94	1.40	0.18	**64.12**
				ε **= 0.20**			
0	**6.00**	5.80	1.20	0.00	-	-	**7.00**
0.5	**7.70**	7.34	1.98	0.06	-	-	**9.38**
0.6	**8.50**	8.10	2.46	0.20	-	-	**10.76**
0.7	**9.76**	9.32	3.10	0.30	-	-	**12.72**
0.8	**12.70**	11.98	5.04	0.56	-	-	**17.58**
0.9	**20.42**	18.84	9.16	1.54	-	-	**29.54**
0.95	**32.48**	27.80	13.86	2.00	-	-	**43.66**
0.99	**45.84**	38.10	16.68	2.22	-	-	**57.00**

(b) cor_u = 0, het_u = 0

α	Sequential	Bai and Perron procedure					
	Total	**1 break**	**2 breaks**	**3 breaks**	**4 breaks**	**5 breaks**	**Total**
				ε **= 0.15**			
0	**5.28**	5.14	1.36	0.10	0.00	0.00	**6.60**
0.5	**7.40**	7.06	2.12	0.28	0.10	0.00	**9.56**
0.6	**8.28**	7.78	2.86	0.46	0.16	0.00	**11.26**
0.7	**9.28**	8.70	4.06	0.80	0.30	0.02	**13.88**
0.8	**12.14**	11.10	6.58	1.60	0.66	0.06	**20.00**
0.9	**20.11**	17.88	12.36	3.74	1.14	0.08	**35.20**
0.95	**30.72**	26.00	17.28	5.40	1.44	0.16	**50.28**
0.99	**45.40**	36.24	20.10	5.20	1.52	0.26	**63.32**

Note: These results are rejection rates (%) based on 5000 Monte Carlo simulations and the data generating process $y_t = \alpha y_{t-1} + \varepsilon_t$ with $\varepsilon_t = iidN(0,1)$. The sample size is T=125 observations. The asymptotic critical values for the nominal 5% level are from Bai and Perron (2003b). We use the parametric correction for serial correlation**: (a)** we allow for heterogeneous variance of the residuals across segments and **(b)** we impose homogeneous distributions of the residuals.

Table 2. Size of the sequential method and the *Bai and Perron procedure*: the non-trending non-parametric case

Estimated model: $y_t = c_j + u_t$

(a) cor_u = 1, het_u = 1

α	Sequential	Bai and Perron procedure					
	Total	1break	2 breaks	3 breaks	4 breaks	5 breaks	Total
				ε = 0.15			
0	**8.76**	8.44	2.80	0.34	0.08	0.00	**11.66**
0.5	**12.12**	11.16	5.64	0.86	0.32	0.02	**18.00**
0.6	**13.56**	12.50	6.92	1.28	0.48	0.04	**21.22**
0.7	**15.62**	14.26	8.94	1.78	0.94	0.08	**26.00**
0.8	**19.92**	17.58	12.54	2.78	1.36	0.22	**34.48**
0.9	**27.88**	23.32	17.16	5.40	2.52	0.34	**48.74**
0.95	**36.64**	29.86	18.92	6.38	2.80	0.48	**58.44**
0.99	**48.12**	33.84	23.36	8.08	2.74	0.64	**68.66**
				ε = 0.20			
0	**7.14**	6.90	1.52	0.04	-	-	**8.46**
0.5	**9.48**	8.98	2.44	0.14	-	-	**11.56**
0.6	**10.42**	9.78	3.30	0.10	-	-	**13.18**
0.7	**11.98**	11.22	4.26	0.48	-	-	**15.96**
0.8	**16.86**	14.58	6.78	0.82	-	-	**22.18**
0.9	**24.12**	21.14	11.66	1.68	-	-	**34.48**
0.95	**33.24**	29.26	13.82	2.24	-	-	**45.32**
0.99	**44.94**	36.54	17.94	3.08	-	-	**57.56**

(b) cor_u = 1, het_u = 0

α	Sequential	Bai and Perron procedure					
	Total	1 break	2 breaks	3 breaks	4 breaks	5 breaks	Total
				ε = 0.15			
0	**6.84**	6.44	1.92	0.10	0.06	0.00	**8.52**
0.5	**7.92**	7.34	3.48	0.64	0.28	0.00	**11.74**
0.6	**8.92**	8.20	4.26	1.14	0.46	0.04	**14.10**
0.7	**10.34**	9.34	5.80	1.52	0.78	0.12	**17.56**
0.8	**13.38**	11.76	9.46	3.22	1.42	0.22	**26.08**
0.9	**21.42**	17.62	15.62	6.46	3.22	0.42	**43.34**
0.95	**31.68**	25.54	19.40	7.84	3.42	0.66	**56.86**
0.99	**49.02**	33.00	24.64	9.66	3.14	0.64	**71.08**
				ε = 0.20			
0	**6.58**	6.34	1.24	0.00	-	-	**7.58**
0.5	**7.40**	7.08	1.92	0.14	-	-	**9.14**
0.6	**8.56**	7.94	2.52	0.16	-	-	**10.62**
0.7	**9.84**	9.30	3.44	0.36	-	-	**13.10**
0.8	**13.00**	11.98	6.04	0.80	-	-	**18.82**
0.9	**21.30**	18.62	11.48	1.96	-	-	**32.06**
0.95	**31.50**	27.34	15.26	2.66	-	-	**45.26**
0.99	**48.50**	37.94	20.94	3.78	-	-	**62.66**

Note: These results are rejection rates (%), based on 5000 Monte Carlo simulations and data generating process $y_t = \alpha y_{t-1} + \varepsilon_t$ with $\varepsilon_t = iidN(0,1)$. The sample size is T=125 observations. The asymptotic critical values for the nominal 5% level are from Bai and Perron (2003b). We use the non-parametric correction for serial correlation**: (a)** we allow for heterogeneous variance of the residuals across segments and **(b)** we impose homogeneous distributions of the residuals

Table 3. Size of the sequential method and the *Bai and Perron procedure*: the trending parametric case

Estimated model: $y_t = c_j + \rho y_{t-1} + \beta t + u_t$

(a) cor_u = 0, het_u = 1

α	Sequential	Bai and Perron procedure					
	Total	1break	2 breaks	3 breaks	4 breaks	5 breaks	Total
			ε = 0.15				
0	**10.96**	10.22	2.66	0.90	0.28	0.02	**14.08**
0.5	**15.30**	13.98	4.58	1.94	0.62	0.16	**21.28**
0.6	**17.66**	16.02	5.98	2.16	0.78	0.22	**25.16**
0.7	**21.30**	18.64	8.08	3.20	1.18	0.18	**31.28**
0.8	**28.34**	23.74	11.94	4.42	1.68	0.44	**42.22**
0.9	**42.54**	32.02	19.56	5.32	2.08	0.30	**59.28**
0.95	**55.02**	39.38	22.68	5.72	2.18	0.22	**70.18**
0.99	**61.46**	44.54	21.88	5.42	1.72	0.22	**73.78**
			ε = 0.20				
0	**11.00**	10.60	1.68	0.32	-	-	**12.60**
0.5	**15.56**	14.74	3.40	0.52	-	-	**18.66**
0.6	**19.10**	17.60	4.20	0.70	-	-	**22.50**
0.7	**21.16**	19.40	5.50	1.12	-	-	**26.02**
0.8	**27.98**	24.52	9.16	1.26	-	-	**34.94**
0.9	**41.50**	33.86	14.30	1.68	-	-	**49.84**
0.95	**52.50**	41.68	17.42	1.44	-	-	**60.54**
0.99	**58.72**	46.78	16.94	1.42	-	-	**65.14**

(b) cor_u = 0,het_u = 0

α	Sequential	Bai and Perron procedure					
	Total	1 break	2 breaks	3 breaks	4 breaks	5 breaks	Total
			ε = 0.15				
0	**10.33**	9.76	2.56	0.90	0.26	0.02	**13.50**
0.5	**15.12**	13.88	4.56	1.94	0.64	0.18	**21.20**
0.6	**17.12**	15.62	5.72	2.18	0.84	0.22	**24.58**
0.7	**21.10**	18.40	7.78	3.04	1.22	0.26	**30.70**
0.8	**28.36**	23.50	11.86	4.42	1.62	0.44	**41.84**
0.9	**41.70**	32.22	191.8	5.32	2.08	0.24	**59.04**
0.95	**54.68**	39.26	22.74	5.62	2.14	0.24	**70.00**
0.99	**61.34**	44.18	21.98	5.44	1.78	0.22	**73.60**

Note: These results are based on 5000 Monte Carlo simulations and data generating process $y_t = \alpha y_{t-1} + \beta t + \varepsilon_t$ with $\varepsilon_t = iidN(0,1)$. The sample size is T=125 observations. The asymptotic critical values for the nominal 5% level are from Bai and Perron (2003b). We use the parametric correction for serial correlation**.** **(a)** we allow for heterogeneous variance of the residuals across segments and **(b)** we impose homogeneous distributions of the residuals.

Table 4. Size of the sequential method and the *Bai and Perron procedure*: the trending non-parametric case

Estimated model: $y_t = c_j + \beta t + u_t$

cor_u = 1, het_u = 0

α	Sequential	Bai and Perron procedure					
	Total	**1break**	**2 breaks**	**3 breaks**	**4 breaks**	**5 breaks**	**Total**
			ε **= 0.15**				
0	**10.88**	10.24	2.72	0.84	0.32	0.08	**14.20**
0.5	**12.46**	11.48	4.86	1.30	0.36	0.08	**18.53**
0.6	**13.94**	12.76	5.62	1.52	0.50	0.06	**21.00**
0.7	**15.72**	14.54	6.22	2.02	0.62	0.12	**24.15**
0.8	**18.88**	17.14	7.80	2.62	0.82	0.12	**28.50**
0.9	**22.26**	20.42	7.94	2.92	1.22	0.20	**32.70**
0.95	**23.84**	21.30	9.34	3.14	1.44	0.28	**35.55**
0.99	**26.42**	23.08	10.30	2.78	1.28	0.22	**37.66**
			ε **= 0.20**				
0	**10.92**	10.60	0.94	0.08	-	-	**11.62**
0.5	**12.50**	11.98	2.00	0.10	-	-	**14.08**
0.6	**13.66**	13.02	2.32	0.16	-	-	**15.50**
0.7	**17.98**	16.98	3.46	0.44	-	-	**20.88**
0.8	**17.98**	16.98	3.46	0.44	-	-	**20.88**
0.9	**20.58**	19.40	4.38	0.56	-	-	**24.34**
0.95	**22.12**	20.50	4.98	0.60	-	-	**26.08**
0.99	**24.76**	22.70	5.64	0.62	-	-	**28.96**

Note: These results are based on 5000 Monte Carlo simulations and data generating process $y_t = \alpha y_{t-1} + \beta t + \varepsilon_t$ with $\varepsilon_t = iidN(0,1)$. The sample size is T=125 observations. The asymptotic critical values for the nominal 5% level are from Bai and Perron (2003b). We use the non-parametric correction for serial correlation**.** We impose homogeneous distributions of the residuals (there is no specification which allows heterogeneous variance of the residuals across segments).

Table 5. Size of the *modified Bai procedure* and the *restricted procedure*

Rejection Rates (%)		**Sequential**	**The modified Bai procedure**					
DGP	α	**Total**	1 break	2 breaks	3 breaks	4 breaks	5 breaks	**Total**
$y_t = \alpha y_{t-1} + \varepsilon_t$ [1]	0.0	**5.0**	4.7	2.2	1.3	1.3	0.9	**10.4**
	0.5	**5.0**	4.9	2.0	1.4	1.3	1.4	**11.0**
	0.6	**5.0**	4.9	2.1	1.6	1.4	1.5	**11.5**
	0.7	**5.0**	4.9	2.0	1.7	1.6	1.3	**11.5**
	0.8	**5.0**	4.9	2.1	1.8	1.7	1.0	**11.5**
	0.9	**5.0**	4.9	2.1	1.9	1.7	0.9	**11.5**
	0.95	**5.0**	4.9	2.3	2.0	1.5	1.0	**11.7**
	0.99	**5.0**	4.9	2.2	1.7	1.5	1.2	**11.5**
$y_t = \alpha y_{t-1} + \beta t + \varepsilon_t$ [2]	0.0	**5.0**	4.7	1.8	1.1	1.1	1.1	**9.8**
	0.5	**5.0**	4.7	1.9	1.5	1.1	0.9	**10.1**
	0.6	**5.0**	4.7	1.8	1.5	1.3	0.8	**10.1**
	0.7	**5.0**	4.7	2.0	1.5	1.3	0.9	**10.4**
	0.8	**5.0**	4.7	1.9	1.5	1.3	0.9	**10.3**
	0.9	**5.0**	4.7	2.2	1.3	1.3	0.9	**10.4**
	0.95	**5.0**	4.9	2.0	1.4	1.3	1.4	**11.0**
	0.99	**5.0**	4.9	2.1	1.6	1.4	1.5	**11.5**

Rejection Rates (%)		**2 breaks (restricted)**	**The restricted procedure**				
DGP	α	**Total**	2 breaks	3 breaks	4 breaks	5 breaks	**Total**
$y_t = \alpha y_{t-1} + \varepsilon_t$ [1]	0.0	**5.0**	3.5	1.8	1.7	1.6	**8.6**
	0.5	**5.0**	3.6	2.0	1.8	1.5	**8.9**
	0.6	**5.0**	3.9	2.2	2.0	1.4	**9.5**
	0.7	**5.0**	4.0	2.3	1.8	1.4	**9.5**
	0.8	**5.0**	4.0	2.5	2.1	1.4	**10.0**
	0.9	**5.0**	4.3	2.4	1.8	1.3	**9.8**
	0.95	**5.0**	4.2	2.4	1.9	1.4	**9.9**
	0.99	**5.0**	4.3	2.5	2.0	1.4	**10.2**
$y_t = \alpha y_{t-1} + \beta t + \varepsilon_t$ [2]	0.0	**5.0**	3.5	1.5	1.7	1.4	**8.1**
	0.5	**5.0**	3.5	1.9	1.8	1.6	**8.8**
	0.6	**5.0**	3.6	1.9	1.8	1.4	**8.7**
	0.7	**5.0**	3.8	2.0	1.7	1.4	**8.9**
	0.8	**5.0**	3.9	2.3	1.7	1.2	**9.1**
	0.9	**5.0**	4.2	2.2	1.5	1.0	**8.9**
	0.95	**5.0**	4.3	2.0	1.7	1.0	**9.0**
	0.99	**5.0**	4.2	2.1	1.3	1.0	**8.6**

Note: These results are based on 5000 Monte Carlo simulations. The sample size is T=125observations with 0.05 trimming. [1] estimate without including a time trend and [2] estimate including a time trend

Table 6. Size adjusted power of the *modified Bai procedure* and the *restricted procedure* on non-trending data including structural change

Rejection Rates (%)		The modified Bai procedure				
DGP	α	1 break	2 breaks	3 breaks	4 breaks	5 breaks
$y_t = \alpha y_{t-1} + \theta DU_t + \varepsilon_t$ $\theta = 0.3$	0.6 0.7 0.8	97.9 93.5 74.0	1.1 1.2 1.2	0.0 0.1 0.6	0.0 0.1 0.3	0.0 0.0 0.5
$y_t = \alpha y_{t-1} + \theta_1 DU_{1t} + \theta_2 DU_{2t} + \varepsilon_t$, $\theta_1 = 0.3, \theta_2 = 0.3$	0.6 0.7 0.8	13.3 3.2 0.7	71.8 59.1 45.9	1.6 2.0 2.5	0.7 1.2 1.4	0.5 1.3 1.3
$y_t = \alpha y_{t-1} + \theta_1 DU_{1t} + \theta_2 DU_{2t} + \varepsilon_t$ $\theta_1 = 0.3, \theta_2 = -0.3$	0.6 0.7 0.8	0.0 0.1 0.0	80.8 72.1 59.7	4.7 3.6 2.5	2.6 2.1 1.4	2.4 2.2 1.7
$y_t = \alpha y_{t-1} + \theta_1 DU_{1t} + \theta_2 DU_{2t} + \theta_3 DU_{3t} + \theta_4 DU_{4t} + \varepsilon_t$ case 1: $\theta_1 = 0.3, \theta_2 = -0.3, \theta_3 = 0.3, \theta_4 = -0.3$	0.6 0.7 0.8	0.0 0.0 0.0	0.5 0.1 0.2	0.8 0.5 0.5	47.1 37.4 33.7	5.3 4.1 3.9
$y_t = \alpha y_{t-1} + \theta_1 DU_{1t} + \theta_2 DU_{2t} + \theta_3 DU_{3t} + \theta_4 DU_{4t} + \varepsilon_t$ case 2: $\theta_1 = 0.3, \theta_2 = 0.3, \theta_3 = -0.3, \theta_4 = -0.3$	0.6 0.7 0.8	0.0 0.0 0.0	0.8 2.9 1.7	2.4 1.2 0.6	55.8 42.2 35.2	5.1 4.5 3.9

Rejection Rates (%)		The restricted procedure			
DGP	α	2 breaks	3 breaks	4 breaks	5 breaks
$y_t = \alpha y_{t-1} + \theta DU_t + \varepsilon_t$ $\theta = 0.3$	0.6 0.7 0.8	25.4 13.4 3.5	3.9 2.4 1.5	0.2 0.1 0.1	0.5 0.1 0.1
$y_t = \alpha y_{t-1} + \theta_1 DU_{1t} + \theta_2 DU_{2t} + \varepsilon_t$, $\theta_1 = 0.3, \theta_2 = 0.3$	0.6 0.7 0.8	0.0 0.0 0.0	0.0 0.0 0.0	0.0 0.0 0.0	0.0 0.0 0.0
$y_t = \alpha y_{t-1} + \theta_1 DU_{1t} + \theta_2 DU_{2t} + \varepsilon_t$ $\theta_1 = 0.3, \theta_2 = -0.3$	0.6 0.7 0.8	78.2 74.2 65.9	11.1 9.9 8.4	2.4 1.8 1.7	4.3 3.7 3.7
$y_t = \alpha y_{t-1} + \theta_1 DU_{1t} + \theta_2 DU_{2t} + \theta_3 DU_{3t} + \theta_4 DU_{4t} + \varepsilon_t$ case 1: $\theta_1 = 0.3, \theta_2 = -0.3, \theta_3 = 0.3, \theta_4 = -0.3$	0.6 0.7 0.8	0.4 0.2 0.2	1.3 1.0 0.8	56.8 48.5 49.7	8.4 7.5 7.7
$y_t = \alpha y_{t-1} + \theta_1 DU_{1t} + \theta_2 DU_{2t} + \theta_3 DU_{3t} + \theta_4 DU_{4t} + \varepsilon_t$ case 2: $\theta_1 = 0.3, \theta_2 = 0.3, \theta_3 = -0.3, \theta_4 = -0.3$	0.6 0.7 0.8	10.5 4.3 4.4	3.5 1.7 1.2	59.3 53.7 50.3	8.5 8.2 10.0

Note: These results are based on 5000 Monte Carlo simulations. The sample size is T=125 observations with 5% trimming.

Table 7. Size adjusted power of the *modified Bai procedure* and the *restricted procedure* on trending data including structural change

Rejection Rates (%)		The modified Bai procedure				
DGP	α	1 break	2 breaks	3 breaks	4 breaks	5 breaks
$y_t = \alpha y_{t-1} + (\beta t) + \theta DU_t + \varepsilon_t$, $\theta = 0.3$	0.6	58.7	4.8	1.5	0.9	0.6
	0.7	53.8	4.5	1.4	0.9	0.9
	0.8	45.6	3.9	1.8	1.3	1.1
$y_t = \alpha y_{t-1} + (\beta t) + \theta_1 DU_{1t} + \theta_2 DU_{2t} + \varepsilon_t$, $\theta_1 = 0.3, \theta_2 = 0.3$	0.6	3.9	32.2	6.4	4.5	4.5
	0.7	3.1	30.3	5.4	4.0	3.9
	0.8	2.3	26.6	4.8	3.2	3.5
$y_t = \alpha y_{t-1} + (\beta t) + \theta_1 DU_{1t} + \theta_2 DU_{2t} + \varepsilon_t$, $\theta_1 = 0.3, \theta_2 = -0.3$	0.6	65.0	26.2	2.0	0.5	0.3
	0.7	55.3	20.5	2.3	1.2	1.3
	0.8	32.3	15.8	3.5	2.3	3.0
$y_t = \alpha y_{t-1} + (\beta t) + \theta_1 DU_{1t} + \theta_2 DU_{2t} + \theta_3 DU_{3t} + \theta_4 DU_{4t} + \varepsilon_t$, case 1: $\theta_1 = 0.3, \theta_2 = -0.3, \theta_3 = 0.3, \theta_4 = -0.3$	0.6	0.6	36.2	3.8	15.8	6.5
	0.7	0.2	25.1	3.0	14.5	5.6
	0.8	0.0	21.1	2.3	11.9	5.8
$y_t = \alpha y_{t-1} + (\beta t) + \theta_1 DU_{1t} + \theta_2 DU_{2t} + \theta_3 DU_{3t} + \theta_4 DU_{4t} + \varepsilon_t$, case 2: $\theta_1 = 0.3, \theta_2 = 0.3, \theta_3 = -0.3, \theta_4 = -0.3$	0.6	1.8	53.0	5.9	13.3	6.5
	0.7	0.4	30.9	4.5	12.2	6.4
	0.8	0.2	18.3	3.1	10.3	4.9

Rejection Rates (%)		The restricted procedure			
DGP	α	2 breaks	3 breaks	4 breaks	5 breaks
$y_t = \alpha y_{t-1} + (\beta t) + \theta DU_t + \varepsilon_t$, $\theta = 0.3$	0.6	19.0	26.4	2.3	4.8
	0.7	16.6	25.3	2.2	3.8
	0.8	12.8	22.4	1.9	3.7
$y_t = \alpha y_{t-1} + (\beta t) + \theta_1 DU_{1t} + \theta_2 DU_{2t} + \varepsilon_t$, $\theta_1 = 0.3, \theta_2 = 0.3$	0.6	5.2	4.5	10.8	16.3
	0.7	4.4	4.1	8.3	14.3
	0.8	3.9	3.7	6.7	10.6
$y_t = \alpha y_{t-1} + (\beta t) + \theta_1 DU_{1t} + \theta_2 DU_{2t} + \varepsilon_t$, $\theta_1 = 0.3, \theta_2 = -0.3$	0.6	75.7	12.6	2.5	4.3
	0.7	70.6	11.2	1.9	3.7
	0.8	58.7	9.2	1.4	3.5
$y_t = \alpha y_{t-1} + (\beta t) + \theta_1 DU_{1t} + \theta_2 DU_{2t} + \theta_3 DU_{3t} + \theta_4 DU_{4t} + \varepsilon_t$, case 1: $\theta_1 = 0.3, \theta_2 = -0.3, \theta_3 = 0.3, \theta_4 = -0.3$	0.6	0.5	1.5	50.2	8.3
	0.7	0.3	0.8	44.3	7.1
	0.8	0.2	0.7	39.5	6.5
$y_t = \alpha y_{t-1} + (\beta t) + \theta_1 DU_{1t} + \theta_2 DU_{2t} + \theta_3 DU_{3t} + \theta_4 DU_{4t} + \varepsilon_t$, case 2: $\theta_1 = 0.3, \theta_2 = 0.3, \theta_3 = -0.3, \theta_4 = -0.3$	0.6	8.0	3.8	55.1	7.7
	0.7	2.3	1.9	45.3	7.5
	0.8	0.7	0.9	33.4	6.0

Note: These results are based on 5000 Monte Carlo simulations. The sample size is T=125 observations with 5% trimming.

Table 8. Unit root tests on real exchange rates

Augmented Dickey-Fuller test

T	Real exchange rate	$\Delta y_t = \mu + \alpha y_{t-1} + \sum_{i=1}^{k} c_i \Delta y_{t-i} + \varepsilon_t$			$\Delta y_t = \mu + \beta t + \alpha y_{t-1} + \sum_{i=1}^{k} c_i \Delta y_{t-i} + \varepsilon_t$		
		α	k	T-stat on α	α	k	T-stat on α
129	Australia	-0.10261	1	-2.62*	-0.17342	1	-3.66 **
117	Belgium	-0.22246	1	-4.12 ***	-0.30877	1	-5.05 ***
129	Canada	-0.06993	0	-1.62	-0.14277	0	-2.98
119	Denmark	-0.06547	6	-1.24	-0.12240	6	-1.97
119	Germany	-0.09041	1	-2.95 **	-0.11217	1	-3.32*
118	Finland	-0.41589	1	-6.02 ***	-0.44080	1	-6.21 ***
119	France	-0.15629	1	-3.55***	-0.22107	1	-4.16***
119	Italy	-0.24655	2	-4.28 ***	-0.24778	2	-4.27***
114	Japan	-0.01636	1	-1.02	-0.08050	7	-1.98
129	Netherlands	-0.09555	1	-2.79 *	-0.11675	1	-3.2 *
129	Norway	-0.12918	1	-3.67 ***	-0.15054	1	-3.96 **
109	Portugal	-0.11736	5	-2.25	-0.12920	5	-2.15
119	Spain	-0.12525	1	-3.24**	-0.12933	1	-3.23 *
119	Sweden	-0.17148	1	-3.72 ***	-0.25125	1	-4.52 ***
107	Switzerland	-0.04288	2	-1.5	-0.12977	2	-2.78
129	United Kingdom	-0.14768	4	-2.61 *	-0.17255	4	-2.74

* significance at the 10% level, ** significance at the 5% level, *** denotes at the 1% level
critical values: model without a trend: -2.58 (10%), -2.89 (5%), -3.51 (1%)
model with a trend: -3.15 (10%), -3.45 (5%), -4.04 (1%).

(T)PPP restricted structural change test

T	Real exchange rate	PPP restricted structural change					TPPP restricted structural change				
		α	Year 1	Year 2	k	τ	α	Year 1	Year 2	k	τ
129	Australia	-0.198	1882	1913	1	-4.10	-0.242	1913	1947	1	5.40*
117	Belgium	-0.292	1922	1930	1	-4.85*	-0.702	1910	1971	7	8.55***
129	Canada	-0.133	1884	1948	1	-2.61	-0.498	1895	1982	8	4.91
119	Denmark	-0.144	1891	1982	1	-3.16	-0.424	1944	1966	3	6.13***
119	Germany	-0.154	1911	1980	8	-4.04	-0.185	1930	1943	1	4.71
118	Finland	-0.507	1916	1974	1	-10.24***	-0.560	1926	1974	1	11.60***
119	France	-0.336	1914	1982	1	-6.42***	-0.388	1914	1982	1	6.63***
119	Italy	-0.372	1919	1942	2	-7.06***	-0.372	1919	1942	2	7.06***
114	Japan	-0.060	1962	1985	7	-2.42	-0.313	1927	1972	1	5.83**
129	Netherlands	-0.196	1882	1968	1	-4.34	-0.230	1944	1968	1	4.88
129	Norway	-0.173	1915	1971	1	-5.05**	-0.238	1915	1967	1	6.24***
109	Portugal	-0.343	1916	1986	1	-6.04***	-0.345	1916	1986	1	5.92**
119	Spain	-0.241	1916	1968	1	-5.85***	-0.242	1916	1968	1	5.74**
119	Sweden	-0.230	1968	1977	1	-4.67*	-0.327	1916	1977	1	5.87**
107	Switzerland	-0.089	1968	1987	1	-2.86	-0.349	1943	1970	1	6.12***
129	United Kingdom	-0.578	1944	1972	3	-7.08***	-0.588	1944	1972	3	7.07***

* significance at the 10% level, ** significance at the 5% level, *** denotes at the 1% level
critical values: model without a trend: -4.61(10%), -4.93(5%), -5.48(1%)
model with a trend: -5.34(10%), -5.59(5%), -6.12(1%)

Table 9. Bai and Perron multiple structural change tests on real exchange rates

Non-parametric		No-trend		Trend		
Country	Time span	Break dates	θ	Break dates	θ	β
Australia	1870-1998	1919	-0.29	-	-	-
Belgium	1880-1996	-	-	-	-	-
Denmark	1880-1998	1972	0.42	-	-	-
Germany	1880-1998	-	-	-	-	-
Finland	1881-1998	-	-	-	-	-
France	1880-1998	1917	-0.28	-	-	-
Japan	1885-1998	1971	1.12	-	-	-
Italy	1880-1998	-	-	-	-	-
Netherlands	1870-1998	1972	0.40	1947, 1970	-0.21, 0.51	0.001
Norway	1870-1998	-	-	1920	0.47	0.008
Portugal	1890-1998	1919	-0.43	1919	-0.63	0.004
Spain	1880-1998	-	-	1920, 1946	-0.60, -0.57	0.012
Sweden	1880-1998	-	-	-	-	-
Switzerland	1892-1998	1972	0.65	1947, 1972	-0.31, 0.41	0.009
United Kingdom	1870-1998	-	-	-	-	-

Parametric		No-trend			Trend			
Country	Time span	Break dates	θ	ρ	Break dates	θ	ρ	β
Australia	1870-1998	1916	-0.07	0.78	-	-	-	-
Belgium	1880-1996	-	-	-	1918, 1979	-0.440, -0.230	0.62	0.012
Denmark	1880-1998	1970	0.11	0.75	1914, 1944	-0.13, -0.14	0.74	0.004
Germany	1880-1998	-	-	-	1920, 1940	0.04, -0.06	0.91	0.001
Finland	1881-1998	-	-	-	1918	-0.20	0.65	0.003
France	1880-1998	1916	-0.07	0.74	1970	0.11	0.70	-0.002
Japan	1885-1998	-	-	-	1923	-0.13	0.83	0.004
Italy	1880-1998	-	-	-	-	-	-	-
Netherlands	1870-1998	-	-	-	-	-	-	-
Norway	1870-1998	-	-	-	1918	-0.14	0.82	0.002
Portugal	1890-1998	1918	-0.10	0.78	1919, 1947	-0.38, -0.27	0.54	0.007
Spain	1880-1998	-	-	-	1918, 1945	-0.21, -0.17	0.75	-
Sweden	1880-1998	-	-	-	-	-	0.69	0.006
Switzerland	1892-1998	1970	0.13	0.82	-	-	0.77	0.008
United Kingdom	1870-1998	-	-	-	1945	-0.09	0.74	0.001

Note: We consider in all cases het_u = 1, except in the non-parametric trend case, where het_u = 0.

Table 10. The modified Bai procedure on real exchange rates

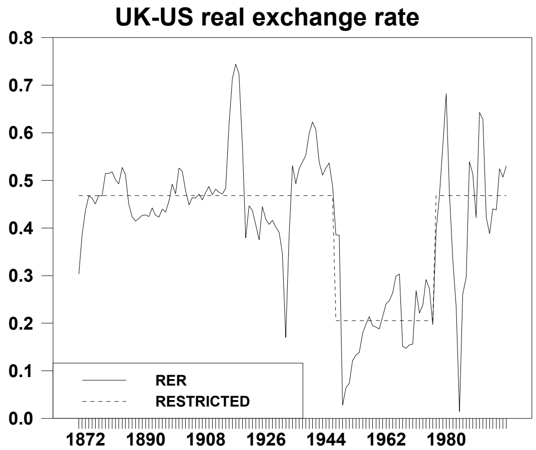
		$y_t = c + \sum_{l=1}^{m} \theta_\ell DU_{\ell t} + \sum_{i=1}^{k} \rho_i y_{t-i} + u_t$					$y_t = c + (\beta t) + \sum_{l=1}^{m} \theta_\ell DU_{\ell t} + \sum_{i=1}^{k} \rho_i y_{t-i} + u_t$					
Country	Hypothesis testing	Statistics	Critical values	Breaks dates	θ	ρ	Statistic	Critical values	Break dates	θ	ρ	β
Australia	0 to 1	13.50	14.26	-	-	-	7.89	15.27	-	-	-	-
	1 to 2	1.78	16.91				5.53	15.02				
Belgium	0 to 1	8.19	11.19	-	-	-	25.60*	12.73	**1912**	0.36	0.55	0.006
	1 to 2	12.32	14.24				23.95*	12.76	**1918**	- 0.61		
Denmark	0 to 1	12.24	13.91	-	-	-	8.80	15.54	-	-	-	-
	1 to 2	3.99	16.70				7.16	15.18				
Finland	0 to1	4.66	10.19	**1912**	0.30	0.55	12.80*	12.13	**1912**	0.27	0.52	0.002
	1 to 2	16.45*	13.00	**1918**	-0.28		17.42*	11.89	**1918**	-0.38		
France	0 to 1	16.36*	11.65	**1916**	-0.09	0.67	17.39*	13.54	**1972**	0.14	0.61	-0.002
	1 to 2	8.24	15.13				9.21	13.56				
Germany	0 to 1	6.39	12.67	-	-	-	5.81	14.63	-	-	-	-
	1 to 2	5.90	16.27				6.22	14.39				
Italy	0 to 1	1.94	11.38	**1939**	0.29	0.72	5.91*	13.81	**1939**	0.27	0.73	0.000
	1 to 2	21.76*	14.58	**1945**	-0.28		17.28*	13.91	**1945**	-0.29		
Japan	0 to1	9.20	17.13	-	-	-	14.53	15.73	-	-	-	-
	1 to 2	5.94	18.53				9.25	15.54				
Netherlands	0 to1	14.26*	13.08	**1970**	0.09	0.78	13.41	15.11	-	-	-	-
	1 to 2	5.89	16.68				4.23	15.11				
Norway	0 to 1	9.41	11.85	-	-	-	14.47*	14.13	**1918**	-0.12	0.80	0.002
	1 to 2	4.45	13.70				12.91	14.31				
Portugal	0 to 1	7.87	13.33	-	-	-	18.50*	14.78	**1918**	-0.35	0.56	0.006
	1 to 2	12.23	16.28				23.59*	14.76	**1947**	-0.25		
Spain	0 to 1	6.14	12.09	-	-	-	13.32	14.81	**1918**	-0.21	0.70	0.004
		10.42	15.71				18.87*	14.86	**1945**	-0.19		
Sweden	0 to 1	6.81	11.56	-	-	-	4.46	13.18	-	-	-	-
		4.20	14.82				8.03	13.23				
Switzerland	0 to 1	16.76*	15.01	**1970**	0.13	0.80	14.65*	13.95	**1972**	0.14	0.72	0.001
	1 to 2	5.07	17.65				8.99	14.07				
UK	0 to 1	5.82	12.69	**1946**	-0.08	0.68	11.22	15.07	-	-	-	-
	1 to 2	17.38*	15.70	**1984**	0.10		10.83	14.76				

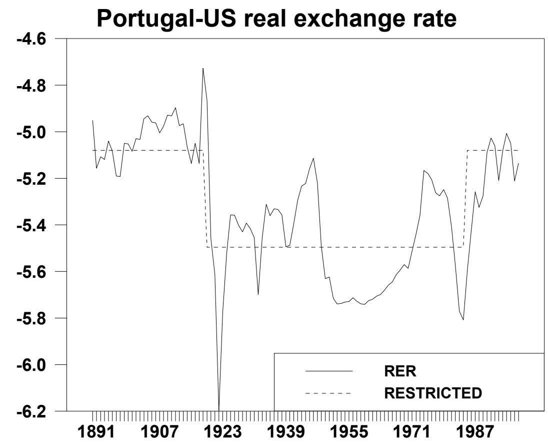
Table 11. The restricted procedure on real exchange rates

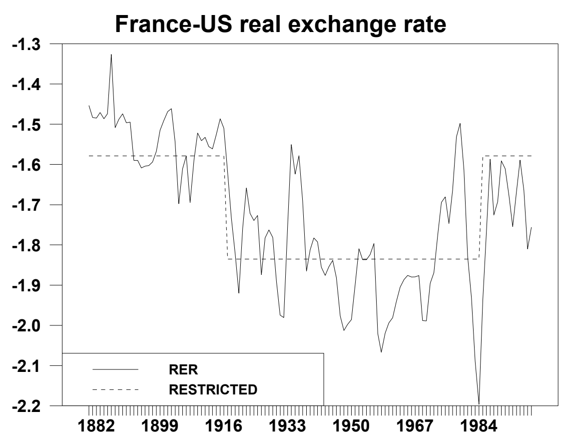
	$y_t = c + \sum_{l=1}^{m} \theta_\ell DU_{\ell t} + \sum_{i=1}^{k} \rho_i y_{t-i} + u_t, \quad \sum_{\ell=1}^{m} \theta_\ell = 0$					$y_t = c + (\beta t) + \sum_{l=1}^{m} \theta_\ell DU_{\ell t} + \sum_{i=1}^{k} \rho_i y_{t-i} + u_t, \quad \sum_{\ell=1}^{m} \theta_\ell = 0$					
Country	Statistics	Critical values	Breaks dates	θ	ρ	Statistic	Critical values	Break dates	θ	ρ	β
Australia	-	-	-	-	-	-	-	-	-	-	-
Belgium	21.33*	10.10	**1918**	-0.24	0.71	32.10*	19.05	**1912**	0.43	0.60	0.003
			1932	0.24				**1918**	-0.43		
Denmark	-	-	**-**	**-**	-	-	-	-	-	-	-
Finland	21.12*	17.19	**1912**	0.29	0.57	26.33*	18.20	**1912**	0.32	0.52	0.001
			1918	-0.29				**1918**	-0.32		
France	22.06*	19.55	**1916**	-0.09	0.69	23.73*	20.10	**1916**	-0.09	0.56	-0.001
			1984	0.09				**1972**	0.09		
Germany	-	-	-	-	-		-	-	-	-	-
Italy	24.04*	18.94	**1939**	0.29	0.72	23.85*	20.48	**1939**	0.29	0.73	0.000
			1945	-0.29				**1945**	-0.29		
Japan	-	-	-	-	-	23.03*	22.95	**1929**	-0.14	0.69	0.005
								1970	0.14		
Netherlands	-	-	-	-	-	-	-	-	-	-	-
Norway	-	-	-	-	-	-	-	**-**	-	-	-
Portugal	20.60*	20.32	**1919**	-0.15	0.70	24.36*	21.71	**1918**	-0.15	0.68	-0.000
			1984	0.15				**1984**	0.15		
Spain	-	-	-	-	-	-	-	-	-	-	-
Sweden	-	-	-	-	-	-	-	-	-	-	-
Switzerland	-	-	-	-	-	23.84*	20.68	**1945**	-0.13	0.64	0.003
								1972	0.13		
UK	23.78	19.73	**1946**	-0.11	0.63	22.94*	21.68	**1946**	-0.10	0.67	0.000
			1976	0.11				**1984**	0.10		

Figure 1. Evidence of PPP or TPPP restricted structural change

A. PPP restricted structural change







B. TPPP restricted structural change