



UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

WORKING PAPERS IN ECONOMICS

No 384

**Testing for Unit Roots in Panel Time Series Models with
Multiple Breaks**

Joakim Westerlund

September 2009

ISSN 1403-2473 (print)
ISSN 1403-2465 (online)

TESTING FOR UNIT ROOTS IN PANEL TIME SERIES MODELS WITH MULTIPLE BREAKS*

Joakim Westerlund[†]
University of Gothenburg
Sweden

September 25, 2009

Abstract

This paper proposes two new unit root tests that are appropriate in the presence of an unknown number of structural breaks. One is based on a single time series and the other is based on a panel of multiple series. For the estimation of the number of breaks and their locations, a simple procedure based on outlier detection is proposed. The limiting distributions of the tests are derived and evaluated in small samples using simulation experiments. The implementation of the tests is illustrated using as an example purchasing power parity.

JEL Classification: C12; C15; C22; F31.

Keywords: Unit root test; Structural break; Outlier detection; Common factor; Purchasing power parity.

1 Introduction

During the last decade, a great deal of research has focused on the search for the best way to characterize or model the dynamic properties of macroeconomic and financial time series. Specifically, the distinction between unit root and stationary processes has become a dominant topic in time series econometrics. Due to its far-reaching economical implications, it

*A previous version of this paper was presented at a seminar at Lund University. The author would like to thank seminar participants and in particular Anindya Banerjee, David Edgerton, Rolf Larsson, Chris Orme and two anonymous referees for many valuable comments and suggestions. The author would also like to thank the Maastricht Research School of Economics of Technology and Organizations for its hospitality during a visit at the Department of Quantitative Economics at the University of Maastricht, where a part of this paper was written. Thank you also to the Jan Wallander and Tom Hedelius Foundation for financial support under research grant number W2006-0068:1.

[†]Department of Economics, University of Gothenburg, P. O. Box 640, SE-405 30 Gothenburg, Sweden. Telephone: +46 31 786 5251, Fax: +46 31 786 1043, E-mail address: joakim.westerlund@economics.gu.se.

has also become a central issue in empirical research, where it has been concluded that many time series can be characterized as unit root processes. The perhaps most canonical example being purchasing power parity (PPP), where the non-stationarity of the real exchange rate has been frequently documented, see for example Choi (2004) for a review of the recent literature.

An important feature common to most unit root studies of this kind is the assumption of parameter stability, or no structural change, and there has been only a few recent attempts to relax it. Yet, ever since the seminal article of Perron (1989), researchers have been well aware of the potential hazards of falsely imposing parameter stability when testing for a unit root. Indeed, as Perron (1989) showed, the ability to reject the unit root null can decrease substantially when the stationary alternative is true but existing structural breaks are ignored.

This is important because the way in which traditional unit root testing is carried out typically involves employing time series that span extended periods of time, which obviously increases the probability of a structural break. The implication is that the inability of many empirical studies to reject the unit root null may well be due to an erroneous omission of structural breaks.

Amsler and Lee (1995) remedy this critic by developing a test based on the Lagrange multiplier principle. The test, which builds on earlier work by Schmidt and Phillips (1992), relaxes the assumption of parameter stability by assuming that the level of the series suffers from a single exogenous, known, break. Although not very unique in itself, this allowance is actually also what makes the test distinct. Specifically, as Amsler and Lee (1995) show, in contrast to existing tests, the asymptotic distribution of their test statistic does not depend on the usual nuisance parameter representing the location of the break, which is of course a very convenient property in the sense that the same critical values can be used regardless of the location of the break.

Lee and Strazicich (2003) take issue with the preference of Amsler and Lee (1995) to treat the break as known, and suggest endogenizing their test using the well-known break estimation procedure of Zivot and Andrews (1992). Specifically, a two-break test is proposed, where the breakpoints are estimated at the minimum of the test values from across all possible break dates.

In this paper, we generalize the test of Amsler and Lee (1995) to the case when there is an

unknown number of breaks in the level of the series. To estimate the unknown breakpoints, we consider a new procedure based on outlier detection, which is advantageous for at least two reasons. Firstly, it is simple and computationally less intensive than other procedures such as the one used by Lee and Strazicich (2003). Secondly, since the estimated breakpoints are valid under both the unit root null and stationary alternative, this makes the outcome of the test easy to interpret.

Moreover, given the potential loss of power that comes from ignoring existing breaks in single time series, it is logical to expect a similar effect when testing for a unit root using a panel of multiple time series. We therefore also propose a panel version of our new test, which can be viewed as a generalization of the exogenous one-break panel test studied by Im *et al.* (2005).

The limiting distributions of the tests are provided and their small-sample properties are investigated through a small simulation study. The results suggest that the asymptotic properties of the tests are borne out well in small samples with small size distortions and good power. This leads us to the conclusion that the new tests should be a valuable addition in applied work. The implementation of the tests is also illustrated empirically using as an example PPP. Using a post-Bretton Woods panel covering 21 countries, we show that in contrast to what many authors have claimed, the failure of PPP cannot be attributed to structural change.

The rest of this paper is organized as follows. Sections 2 and 3 present the model, the new tests and their limiting distributions under the null hypothesis of a unit root with known breaks, which is extended in Section 4 to the case with unknown breaks. Section 5 is concerned with the simulation study, whereas Section 6 contains the empirical application. Section 7 concludes.

2 Model and assumptions

In this section, we generalize the model of Amsler and Lee (1995) in two directions. Firstly, we allow for multiple breaks, and secondly, we consider a more flexible panel setting. Let us therefore consider the panel data variable y_{it} , observable for $t = 1, \dots, T$ time series and $i = 1, \dots, N$ cross-sectional units, whose data generating process can be written in the following

way:

$$y_{it} = \alpha_i + \beta_i t + \delta_i' D_{it} + u_{it}, \quad (1)$$

where D_{it} is an r -dimensional vector of break dummies such that $D_{jit} = 1(t > B_{ji})$, where $1(x)$ is the indicator function and B_{ji} denotes the location of break number j . Thus, in this model, α_i represents the level of the series before any break takes place and δ_{ji} represents the change in the level at the time of break j . The error u_{it} has the following decomposition:

$$u_{it} = \lambda_i' f_t + e_{it}, \quad (2)$$

where the m -dimensional vector of common factors f_t and the loading λ_i represent the common component of u_{it} , while the scalar e_{it} represent the idiosyncratic component. These are assumed to be generated as

$$\Delta f_t = \Pi f_{t-1} + v_t, \quad (3)$$

$$\Delta e_{it} = \rho_i e_{it-1} + w_{it} \quad (4)$$

with w_{it} satisfying $\gamma_i(L)w_{it} = \epsilon_{it}$, where $\gamma_i(L) = 1 - \sum_{s=1}^p \gamma_{is}L^s$ is a polynomial in the lag operator L , and ϵ_{it} is independent across both i and t with mean zero and variance σ_i^2 . Similarly, $\Gamma(L)v_t = \eta_t$, where $\Gamma(L) = I_m - \sum_{s=1}^q \Gamma_s L^s$ with η_t being a mean zero disturbance with covariance matrix Σ that is independent of ϵ_{it} and across t .

The fact that ϵ_{it} is cross-sectionally independent implies that any dependence across units is restricted to the common factors. The extent of this dependence is determined by λ_i . Thus, the factors make the units correlated through u_{it} . In terms of the PPP example discussion in the introduction, it is convenient to think of f_t as comprised of the common base currency and possibly also other factors that are common across the members of the panel.

One way to ensure that ϵ_{it} and η_t are independent is to assume that f_t is strictly exogenous. Apart from this, however, the restrictions placed on f_t are very weak and permit for a wide range of possibilities when it comes to the order of integration of f_t , as determined by the rank of Π . If the rank is full, the elements of f_t are all stationary, whereas if the rank is zero, then they are all non-stationary. If Π has reduced rank, then the elements of f_t are cointegrated. It should be pointed out, however, that the tests that we propose are designed to test for a unit root in e_{it} . Thus, unless Π is a full rank matrix, a rejection of the null hypothesis of a unit root in e_{it} does not necessarily mean that y_{it} is stationary. This means that in

practice f_t must also be tested for unit roots, which can be done by using any conventional unit root test.

For now we also assume that f_t , B_{ji} and r are known. But this is only for simplicity, and can be relaxed at the expense of more complicated proofs, see Sections 4 and 6 for detailed discussions.

3 The unit root test

Having laid out the key assumptions that characterize the model of interest, in this section we consider the problem of testing for a unit root in e_{it} . We begin by developing a time series test that applies to each individual unit, and then we show how these can be combined into a panel test.

3.1 A time series test

Testing for a unit root in e_{it} is equivalent to testing $H_0 : \rho_i = 0$ versus $H_1 : \rho_i < 0$. This restriction can be tested using the Lagrange multiplier, or score, principle, which states that the score must have zero mean when evaluated at the true parameters under the null hypothesis. Suppose therefore that ϵ_{it} is normal, in which case we have the following log-likelihood function:

$$L = \text{constant} - \frac{T-p}{2} \sum_{i=1}^N \ln(\sigma_i^2) + \frac{1}{2} \sum_{i=1}^N \frac{1}{\sigma_i^2} \sum_{t=p+1}^T \epsilon_{it}^2,$$

which can be differentiated with respect to ρ_i and then evaluated at the restricted maximum likelihood estimates, giving

$$\frac{\partial L}{\partial \rho_i} \propto \sum_{t=p+2}^T Q_X \Delta S_{it} Q_X S_{it-1},$$

where Q_X is the least squares residual operator, with

$$Q_X S_{it} = S_{it} - \sum_{t=p+2}^T S_{it-1} X'_{it} \left(\sum_{t=p+2}^T X_{it} X'_{it} \right)^{-1} X_{it}$$

being the residual from regressing Y_{it} onto $X_{it} = (1, \Delta S_{it-1}, \dots, \Delta S_{it-p})'$. Moreover,

$$S_{it} = y_{it} - \hat{\alpha}_i - \hat{\beta}_i t - \hat{\delta}'_i D_{it} - \hat{\lambda}'_i f_t, \quad (5)$$

where $\hat{\alpha}_i = y_{ip+1} - \hat{\beta}_i - \hat{\delta}'_i D_{ip+1} - \hat{\lambda}'_i f_{p+1}$ is the restricted maximum likelihood estimator of α_i under the null. The corresponding maximum likelihood estimates $\hat{\beta}_i$, $\hat{\delta}_i$ and $\hat{\lambda}_i$ of β_i , δ_i and λ_i , respectively, are obtained by applying least squares to

$$\Delta y_{it} = \beta_i + \delta'_i \Delta D_{it} + \lambda'_i \Delta f_t + \Delta e_{it}. \quad (6)$$

The point here is that score is proportional to

$$\tau_{\rho_i} = \frac{\sum_{t=p+2}^T Q_X \Delta S_{it} Q_X S_{it-1}}{\hat{\sigma}_i \sqrt{\sum_{t=p+2}^T (Q_X \Delta S_{it})^2}},$$

the t -statistic for testing $H_0 : \rho_i = 0$ versus $H_1 : \rho_i < 0$ in

$$\Delta S_{it} = \rho_i S_{it-1} + \phi'_i X_{it} + \text{error}. \quad (7)$$

The asymptotic distribution of this test statistic under H_0 can be deduced from Westerlund and Edgerton (2006), who show that as $T \rightarrow \infty$,

$$\begin{aligned} \frac{1}{T} \sum_{t=p+2}^T Q_X \Delta S_{it} Q_X S_{it-1} &= \frac{1}{\gamma_i(1)} \frac{1}{T} \sum_{t=p+2}^T \epsilon_{it} w_{it-1} + O_p \left(\frac{1}{\sqrt{T}} \right) \\ &= -\frac{1}{2\gamma_i(1)} \frac{1}{T} \sum_{t=p+2}^T \epsilon_{it}^2 + O_p \left(\frac{1}{\sqrt{T}} \right) = -\frac{\sigma_i^2}{2\gamma_i(1)} + O_p \left(\frac{1}{\sqrt{T}} \right) \end{aligned}$$

with $w_{it} = \sum_{s=p+2}^t (\epsilon_{is} - \bar{\epsilon}_i)$, and

$$\begin{aligned} \frac{1}{T^2} \sum_{t=p+2}^T (Q_X S_{it-1})^2 &= \frac{1}{\gamma_i(1)^2} \frac{1}{T^2} \sum_{t=p+2}^T (Q_1 w_{it-1})^2 + O_p \left(\frac{1}{\sqrt{T}} \right) \\ &\rightarrow_w \frac{\sigma_i^2}{\gamma_i(1)^2} \int_0^1 (Q_1 V_i(s))^2 ds, \end{aligned}$$

where \rightarrow_w signifies weak convergence, $Q_1 V_i(s) = V_i(s) - \int_0^1 V_i(r) dr$, $V_i(s) = W_i(s) - sW_i(1)$ and $W_i(s)$ is a standard Brownian motion on $s \in [0, 1]$. As usual, we use $y_T = O_p(T^r)$ to signify that y_T is at most order T^r in probability, and $y_T = o_p(T^r)$ in case y_T is of smaller order in probability than T^r .¹ It follows that as $T \rightarrow \infty$,

$$\tau_{\rho_i} = -\frac{\sigma_i}{2\sqrt{\frac{1}{T^2} \sum_{t=p+2}^T (Q_1 w_{it-1})^2}} + O_p \left(\frac{1}{\sqrt{T}} \right) \rightarrow_w -\frac{1}{2\sqrt{\int_0^1 (Q_1 V_i(s))^2 ds}}. \quad (8)$$

There are several things about (8) that are worthy of further discussion. Firstly, the asymptotic distribution is free of nuisance parameters, which means that it is invariant with

¹If y_T is deterministic, then $O_p(T^r)$ and $o_p(T^r)$ are replaced by $O(T^r)$ and $o(T^r)$, respectively.

respect to both the common factors and the breaks. This is of course very convenient because it means that we may use the same critical values as given in Tables 1A and 1B of Schmidt and Phillips (1992) for their non-break test, and proceed as if there were no breaks or cross-section dependence at all. Thus, there is no need to tabulate different critical values for different break structures as in for example Perron (1989), or for different values of r as in Pesaran (2007).

The result is also very unusual. In fact, as far as we are aware this is the only test around that is invariant with respect to both breaks and cross-section dependence. The difference lies with the first-differencing, which means that we can use standard asymptotic arguments for stationary processes to show that the estimated coefficients converge to constants as T grows. This stands in sharp contrast to the conventional Dickey–Fuller approach, which is based on estimating the parameters from (1) using the data in levels. However, since this regression is spurious, the estimated regression parameters do not converge to constants, but in fact remain random even asymptotically. It is this difference that makes the asymptotic distribution of the new test relatively simple.

Secondly, since the breaks are allowed under both the null and alternative hypotheses, there is no confusion about the interpretation of the test outcome. Consider for example the recently proposed unit root test of Kapetanios (2005), which is similar to ours in the sense that it is general enough to allow for more than one break. The problem with this test is that the breaks are only permitted under the stationary alternative. Thus, a rejection of the null does not necessarily imply a rejection of a unit root *per se* but rather a rejection of a unit root without breaks, which calls for careful interpretation of the test result in applied work. In particular, with breaks under the null, researchers might incorrectly conclude that a rejection of the null indicates evidence of stationarity with a break, when in fact the series is non-stationary with breaks.²

Thirdly, because the test is asymptotically similar with respect to the breaks under the null hypothesis, the asymptotic distribution will be unaffected if we were to dispense with the assumption of known breaks, which is likely to be unduly restrictive for most empirical purposes. In fact, as shown by Amsler and Lee (1995), the asymptotic distribution of the

²Similarly, Gadea *et al.* (2004) propose a test for the joint hypothesis of a unit root and no breaks, with which they find evidence of a stationary real exchange rate suggesting that PPP holds. The above discussion suggests that this conclusion may be misleading in the sense that the proposed test cannot discriminate between PPP and a non-stationary real exchange rate with breaks.

test remains the same even if the breaks are misplaced. Hence, the distribution is unaffected even if we employ an inconsistent estimator of the breakpoints. The problem is that incorrect placement or exclusion of the breaks makes the test biased towards accepting the null. Thus, although the breaks do not affect the null distribution, they do affect the test by reducing its power, which is why accounting for them is important.

3.2 A panel data test

Given that the asymptotic null distribution of the individual τ_{ρ_i} statistic is free of nuisance parameters, the various panel unit root tests developed in the literature for the case of cross-sectionally independent errors and no breaks can also be applied to the present more general case.

Our interest lies in testing the null hypothesis that all units are non-stationary versus the alternative that there is at least one unit that is stationary, which can be expressed as $H_0 : \rho_i = 0$ for all i against $H_1 : \rho_i < 0$ for some i . Thus, in terms of our empirical application, we are interested in testing the null of no PPP against the alternative that there is at least some evidence in favor of PPP.³

For testing this hypothesis, we propose using the following normalized cross-sectional average of the individual τ_{ρ_i} statistics for each unit:

$$\tau_\rho = \frac{1}{\sqrt{N}\sigma_\tau} \sum_{i=1}^N (\tau_{\rho_i} - \mu_\tau),$$

where μ_τ and σ_τ^2 are the mean and variance of the limiting distribution in (8), which can be obtained using simulations.⁴ For this purpose, we generate 10,000 random walks of length $T = 1,000$. By using these random walks as a simulated Brownian motions, it is possible to evaluate the expression in (8) and then to compute the moments. The simulated expectation and variance based on this method are -1.969 and 0.323 , respectively.⁵

The asymptotic distribution of τ_ρ under H_0 is readily deduced by using (8), from which

³As pointed out by Choi (2004), a rejection of H_0 can be difficult to interpret since it is not clear for which of the countries of the panel PPP holds. To alleviate this problem, the author suggest a method to sequentially classify each country. However, since interest usually lies in testing whether there is at least some evidence of PPP, a test of H_0 versus H_1 is still informative.

⁴ τ_ρ can be seen as a generalization of the exogenous one-break test studied by Im *et al.* (2005) in the case of cross-sectionally independent errors.

⁵Im *et al.* (2005) use repeated sampling of the test statistic to compute small-sample critical values for different combinations of T and p_i . These are reported in Table 1 of their paper.

it follows that as $N, T \rightarrow \infty$ with $\frac{N}{T} \rightarrow 0$,

$$\begin{aligned}\tau_\rho &= -\frac{1}{\sqrt{N}\sigma_\tau} \sum_{i=1}^N \left[\frac{\sigma_i}{2\sqrt{\frac{1}{T^2} \sum_{t=p+2}^T (Q_1 w_{it-1})^2}} + \mu_\tau \right] + O_p\left(\frac{\sqrt{N}}{\sqrt{T}}\right) \\ &= -\lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}\sigma_\tau} \sum_{i=1}^N \left[\frac{1}{2\sqrt{\int_0^1 (Q_1 V_i(s))^2 ds}} + \mu_\tau \right] \rightarrow_w N(0, 1).\end{aligned}$$

This result shows that the main effect of summing over the cross-sectional dimension is to smooth out the Brownian motion dependency for each unit, leading to an asymptotic normal distribution. The condition that $\frac{N}{T} \rightarrow 0$ as $N, T \rightarrow \infty$ is standard even when testing for unit roots in independent panels. The reason for this is the assumed heterogeneity, whose elimination induces an estimation error in T , which is then aggravated when pooling across N . The condition that $\frac{N}{T} \rightarrow 0$ prevents this error from having a dominating effect upon τ_ρ .

4 Unknown breaks

4.1 Existing procedures

Arguably, the single most popular unit root testing procedure with unknown breaks is that of Zivot and Andrews (1992), in which a single breakpoint can be estimated via grid search at the minimum of the individual unit root test statistics from across all possible breakpoints. However, as pointed out by Kapetanios (2005), extending this one-break grid search to r breaks is clearly computationally extremely demanding and practically infeasible for $r > 3$.⁶ Another drawback of this approach is that r must be known. Thus, it is not possible to test whether the number of breaks in fact is equal to r or not.

Moreover, with an invariant test such as ours, this method may well result in deceptive inference. Consider as an example the LM_τ test of Lee and Strazicich (2003), which in their model A amounts to applying the Zivot and Andrews (1992) procedure to the two-break τ_i test.⁷ The authors argue on page 1083 that:

the asymptotic null distribution of the two-break LM unit root test for model A is invariant to the location and magnitude of structural breaks. This property follows from the results shown in Amsler and Lee (1995) for their exogenous

⁶See Gadea *et al.* (2004) for a good discussion and illustration of this issue in the context of PPP.

⁷In Lee and Strazicich (2003) model A refers to the case with two breaks in the constant but no breaks in the trend, which is the same as (1) with $r = 2$.

one-break LM unit root test. Fortunately, this same outcome carries over to the endogenous break LM unit root test. Thus, the asymptotic distribution of the endogenous break LM unit root test will not diverge in the presence of breaks under the null and is robust to their misspecification.

A natural interpretation of this discussion is that because the test is invariant, taking the minimum should not affect its asymptotic distribution. The asymptotic critical values should therefore be the same as for the exogenous test of Schmidt and Phillips (1992).

However, this is incorrect. The problem is that by comparing the minimum statistic with the asymptotic critical value for the exogenous test, one is essentially taking an order statistic and treating it as an ordinary statistic, which is likely to result in size distortions. The reason is that because the same critical values are used, LM_τ will tend to be too large in absolute value, thus making it biased towards rejecting the unit root null. This is seen in Table 2 of Lee and Strazicich (2003), where they report small-sample critical values based on repeated sampling of the minimum statistic itself, which are of course correct even if the asymptotic distribution happens to be misspecified.⁸ For example, the 5% critical value reported in Lee and Strazicich (2003) is -3.842 , while in Schmidt and Phillips (1992) it is only -3.02 . Thus, the asymptotic distribution for the endogenous, minimum, test is not the same as the one that applies to the exogenous test.

But even if one uses the small-sample critical values, the approach of Lee and Strazicich (2003) is still flawed in the sense that the precision of the estimated breakpoints is likely to be very poor. In fact, it is not difficult to see that since the asymptotic distribution in Theorem 1 is independent of B_{ji} , the breakpoint estimator is going to be uniformly distributed, even asymptotically. Thus, the precision is not only going to be poor but nonexistent, which is again likely to result in a loss of power.

4.2 A procedure based on outlier detection

Given the above mentioned problems we look for a more feasible approach that does not require the knowledge of r , and that is less likely to result in deceptive inference. One suggestion towards this end is to treat the estimation problem as a model selection issue, and to estimate the breakpoints at the minimum of the sum of squared residuals obtained from

⁸Thus, while their discussion of the asymptotic distribution is potentially misleading, being based on repeated sampling of the minimum statistic, the small-sample critical values provided by Lee and Strazicich (2003) are still correct.

the test regression in (7). This route is taken by Kapetanios (2005), who proposes a vary efficient grid search scheme that requires only $O(T)$ least squares operations for any r . Unfortunately, this procedure only guaranties consistent estimates of the break fractions under the stationary alternative, and it may therefore be worthwhile to seek other alternatives.⁹

An even simpler approach, that is perhaps more natural in our case, is to treat this as an outlier detection problem, and to estimate the breaks from the first differenced regression in (6), which in contrast to (7) is always stationary. The standard approach to do this is to estimate (6) and to construct a t -test for the presence of an outlier. Such a test is constructed for all possible dates and the maximum is taken. This value is then compared with some critical value to decide if an outlier is present.

The procedure that we propose is adapted from Chen and Liu (1993), and can be described as follows. In case of a unit root the model in differences can be written as

$$\gamma_i(L)\Delta y_{it} = \gamma_i(1)\beta_i + \gamma_i(L)\delta'_i\Delta D_{it} + \gamma_i(L)\lambda'_i\Delta f_t + \epsilon_t,$$

which under the null hypothesis of no outlier reduces to

$$\gamma_i(L)\Delta y_{it} = \gamma_i(1)\beta_i + \gamma_i(L)\lambda'_i\Delta f_t + \epsilon_t,$$

or

$$Q_V\Delta y_{it} = Q_V\epsilon_t, \tag{9}$$

where Q_V is the residual operator based on $V_{it} = (1, \Delta y_{it-1}, \dots, \Delta y_{it-p}, \Delta f'_{it}, \dots, \Delta f'_{it-p})'$. Suppose now that we would like to test the null hypothesis of the absence of a single outlier at time b_{1i} , which may or may not be equal to B_{1i} , the true breakpoint. This makes it convenient to rewrite (9) as

$$Q_V\Delta y_{it} = \delta_{1i}\Delta D_{1it} + Q_V\epsilon_t \tag{10}$$

with $\delta_{1i} = 0$ and $\Delta D_{1it} = 1(t = b_{1i})$. The Wald statistic for testing the null hypothesis that $\delta_{1i} = 0$ in this regression is given by

$$w_{\delta_{1i}}(b_{1i}) = \frac{(\sum_{t=p+2}^T \Delta D_{1it} Q_V \Delta y_{it})^2}{\hat{\sigma}_i^4 \sum_{t=p+2}^T (\Delta D_{1it})^2},$$

⁹Sabatè *et al.* (2003) uses a similar approach when testing for PPP. The problem is that in order to estimate the appropriate number of breaks to use in the unit root test, the authors have to assume beforehand that the real exchange rate is in fact stationary.

where

$$\begin{aligned}\sum_{t=p+2}^T \Delta D_{1it} Q_V \Delta y_{it} &= \Delta y_{ib_{1i}} - \sum_{t=p+2}^T \Delta y_{it} V'_{it} \left(\sum_{t=p+2}^T V_{it} V'_{it} \right) V_{ib_i} \\ &= \epsilon_{ib_{1i}} - \sum_{t=p+2}^T \epsilon_{it} V'_{it} \left(\sum_{t=p+2}^T V_{it} V'_{it} \right) V_{ib_i} = \epsilon_{ib_{1i}} + O_p \left(\frac{1}{\sqrt{T}} \right)\end{aligned}$$

and $\sum_{t=p+2}^T (\Delta D_{1it})^2 = 1$, and by similar arguments,

$$\begin{aligned}\hat{\sigma}_i^2 &= \frac{1}{T} \sum_{t=p+2}^T (Q_V \Delta y_{it} - \hat{\delta}_{1i} \Delta D_{1it})^2 = \frac{1}{T} \sum_{t=p+2}^T (Q_V \Delta y_{it})^2 + O_p \left(\frac{1}{\sqrt{T}} \right) \\ &= \frac{1}{T} \sum_{t=p+2}^T (Q_V \epsilon_{it})^2 + O_p \left(\frac{1}{\sqrt{T}} \right) = \frac{1}{T} \sum_{t=p+2}^T \epsilon_{it}^2 + O_p \left(\frac{1}{\sqrt{T}} \right) = \sigma_i^2 + O_p \left(\frac{1}{\sqrt{T}} \right),\end{aligned}$$

where $\hat{\delta}_{1i}$ is the least squares estimate of δ_{1i} . It follows that

$$w_{\delta_{1i}}(b_{1i}) = \frac{\epsilon_{ib_{1i}}^2}{\sigma_i^2} + O_p \left(\frac{1}{\sqrt{T}} \right). \quad (11)$$

where $\epsilon_{ib_{1i}}/\sigma_i$ has mean zero and unit variance. We can define the following test statistic for the null of no outlier:

$$w_{\delta_{1i}} = \max_{b_{1i}} w_{\delta_{1i}}(b_{1i}).$$

While free of the nuisance parameters that characterize the serial and cross-sectional correlation properties of the errors, the asymptotic distribution of $w_{\delta_{1i}}$ is complicated by the fact that the maximum has been taken, which is the same problem as in testing for outliers in stationary time series. The standard practice in the literature is arbitrary and consists of rejecting the no outlier null if the maximal t -statistic is greater than some predetermined constant between three and four, which has no reference to any significance level. As a response to this, in this section we use extreme value theory to obtain the asymptotic critical value of $w_{\delta_{1i}}$.¹⁰

Suppose that ϵ_{it} is normal, and let $w_{\delta_{1i}}^\circ = \max_{b_{1i}} \epsilon_{ib_{1i}}^2 / \sigma_i^2$, $b_T = F^{-1}(1 - 1/T)$ and $a_T = F^{-1}(1 - 1/Te) - b_T$, where $F^{-1}(x)$ is the inverse of the chi-squared distribution function with one degree of freedom. By using the results of Embrechts *et al.* (1997) it can be shown that

$$\frac{w_{\delta_{1i}}^\circ - b_T}{a_T} \rightarrow_d G(x) \quad (12)$$

as $T \rightarrow \infty$, where \rightarrow_d denotes convergence in distribution and $G(x) = \exp(-e^{-x})$ is the

¹⁰See and Chareka *et al.* (2006) for a similar approach.

Gumbel distribution. But $w_{\delta_{1i}} = w_{\delta_{1i}}^\circ + o_p(1)$ from (11), and so we get

$$\begin{aligned} \left| P\left(\frac{w_{\delta_{1i}} - b_T}{a_T} \leq x\right) - P\left(\frac{w_{\delta_{1i}}^\circ - b_T}{a_T} \leq x\right) \right| &\leq P\left(x - \varepsilon \leq \frac{w_{\delta_{1i}}^\circ - b_T}{a_T} \leq x + \varepsilon\right) \\ &+ P\left(\left|\frac{w_{\delta_{1i}} - w_{\delta_{1i}}^\circ}{a_T}\right| > \varepsilon\right) \rightarrow 0 \end{aligned}$$

with $\varepsilon > 0$. Hence, $(w_{\delta_{1i}} - b_T)/a_T \rightarrow_d G(x)$ as $T \rightarrow \infty$, suggesting that the appropriate critical value at significance level α is given by $C_\alpha = -\ln(-\ln(1 - \alpha))$.

Our outlier detection procedure can now be summarized as follows:

1. Regress Δy_{it} onto V_{it} and obtain the associated residuals, $Q_V \Delta y_{it}$.
2. Estimate (10) by regressing the step-1 residuals onto ΔD_{1it} and obtain $w_{\delta_{1i}}(b_{1i})$ for all possible breakpoints $p + 1 < b_{1i} < T - p - 1$.
3. Compute $w_{\delta_{1i}}$ and compare $(w_{\delta_{1i}} - b_T)/a_T$ to the critical value from the Gumbel distribution, C_α .
4. If $(w_{\delta_{1i}} - b_T)/a_T < C_\alpha$, then there are no outliers, and so the procedure is stopped.
5. However, if $(w_{\delta_{1i}} - b_T)/a_T > C_\alpha$, an outlier is detected at date

$$\hat{B}_{1i} = \arg \max_{b_{1i}} w_{\delta_{1i}}(b_{1i}),$$

which is then removed from the data.

6. Once the outlier observation has been removed from the step-1 residuals, (10) is again estimated and tested for an outlier. This continues until the test fails to reject, or until the number of observations become too small.

Note that the procedure estimates jointly both the number of breaks and their locations. Moreover, because we are dealing with differenced series, which are stationary, $w_{\delta_{1i}}$ is asymptotically independent at each step of the iterations. Thus, the same critical values apply at each step. Also, since (12) only requires identically but not necessarily independent errors, the critical values are robust to violations from the unit root restriction used in deriving (9).

Apart from being simple this procedure has the advantage of not requiring any knowledge about r . In fact, r may be zero. If no outlier is detected, the unit root testing is carried

out as described in Section 3 with S_{it} based on no dummy variables, whereas if there is at least one outlier, then S_{it} is augmented with one impulse dummy for each. Each iteration in the procedure identifies one outlier.¹¹ This is of course very different from the Zivot and Andrews (1992) procedure, in which a nonzero value of r has to be stipulated beforehand.

Another advantage with using $w_{\delta_{it}}$ is that it constitutes a consistent test regardless of whether the unit root null is true or not. This again follows from differencing, which means that we are effectively working with stationary series even though their levels may be non-stationary. Of course, there is no claim of consistency of the resulting breakpoint estimator, and we do not prove here the asymptotic properties of \hat{B}_{ji} . However, intuition suggests that this approach should perform well in practice, and our simulation results confirm this. Moreover, because of the invariance of the unit root test, consistency of \hat{B}_{ji} is not a requirement, at least not under the null.

5 Simulation experiments

5.1 Setup

In this section, we investigate briefly the small-sample properties of the new tests by means of simulations using (1)–(4) to generate the data. For simplicity, we assume that $\gamma_i(L) = 1 - \phi L$, $\Gamma(L) = I_m$, $\epsilon_{it} \sim N(0, \sigma_i^2)$ with $\sigma_i^2 \sim U(0.5, 1, 5)$ and $\eta_t \sim N(0, I_m)$, so that w_{it} is heteroskedastic and possibly also autocorrelated but v_t is standard normal. Also, since the order of integration of the common factor did not affect the results, we set $\Pi = -I_m$.

The data are generated with $\lambda_i \sim N(1, 1)$, $\alpha_i = \beta_i = 0$ and a common breakpoint $B_{ji} = B_j$ for all i . As for the number of structural breaks, r , we consider two cases. In the first, $r = 1$ with $B_1 = 0.3T$, while in the second, $r = 2$ with $B_1 = 0.3T$ and $B_2 = 0.7T$. The break coefficients are generated as $\delta_{ji} = \delta_i$ for all i with $\delta_i \sim N(\delta, 1)$. In the size simulations, we set $\rho_i = \rho = 0$, while in the power simulations, $\rho = -0.1$.

All results are based on 2,000 replications, where the first 100 time series observations in each replication is discarded to avoid possible initial value effects. The significance level is set to 5% throughout, and all powers are adjusted for size. Some results from the accuracy of the estimated breakpoints are also reported. All computational work is done in GAUSS.

¹¹ Although the maximum number of iterations, and hence outliers, could in principle be restricted, our simulation results suggest that this is not necessary. In some circumstances it may also make sense to restrict the length of each regime so as to prevent estimation of neighboring breakpoints, see Section 6 for a discussion.

In constructing the unit root tests, the lag length p was selected in the same way as in Vogelsang (1999), namely using a recursive general-to-specific t -test on the last lag with a significance level of 5% starting at a maximum order of $T^{1/3}$. For the outlier detection procedure, the 5% critical value from the Gumbel distribution was used. For the τ_{ρ_i} test, the critical value -3.02 was taken from Table 1A in Schmidt and Phillips (1992). The τ_ρ test is computed using the simulated moments from Section 3 and compared to the 5% critical value -1.645 from the standard normal distribution.

5.2 Size and size-adjusted power

We begin by considering the size of the tests. We are interested in comparing τ_{ρ_i} with τ_ρ , but also in analyzing how these tests perform depending on the treatment of the breaks, which can either be known, estimated or simply ignored. The results are presented in Table 1.

It is seen that within the class of individual and panel tests the ones based on ignoring the breaks generally perform best, which is to be anticipated given that the asymptotic null distribution of τ_{ρ_i} , and therefore also τ_ρ , is independent of the breaks. Hence, even though there are breaks in the data generating process, the no-break tests are actually expected to perform well here. However, the other tests are almost as accurate, and perform only slightly worse. As expected, none of the tests appear to be affected much by the location of the breaks.

A notable exception from the otherwise so good performance is when $r = 2$ and $T = 100$, in which case the panel tests tend to become oversized. This is due to the distortions at the individual level, which, although very small, have a tendency of accumulating and to become more serious as N grows.

The power results reported in Table 2 generally coincide with what might be expected from the asymptotic theory. Firstly, the power increases with the size of the sample, which is presumably a reflection of the consistency of the tests. The panel tests lead to the best performance by far with almost perfect power in a majority of the cases. Thus, there are potentially large power gains to be made by exploring the cross-sectional dimension. Secondly, within the class of individual and panel tests the no-break tests are generally the least powerful, which corroborates the result that erroneous omission of breaks should affect the tests by lowering their power. Thirdly, the tests based on estimated breaks are almost as powerful as the tests based on known breaks, which is of course very good news in appli-

cations, where there is usually little or no *a priori* knowledge about the location of the break and number of factors.

5.3 Break estimation accuracy

By treating the breaks as unknown, the test results can also be evaluated in terms of the accuracy of the estimated breaks. Table 3 therefore presents the correct selection frequencies for both the number of breaks and their locations.

The accuracy does not appear to be affected by the value of ρ , which is consistent with our outlier detection procedure being valid under both the unit root null and stationary alternative. We similarly see that the performance is basically the same for the two values of γ considered. On the other hand, the accuracy depends to a large extent on both the number of breaks and their magnitude, which is to be expected because a larger number of smaller breaks are more difficult to detect. With reasonably sized breaks, however, accuracy is generally good.¹²

The good precision of the estimated breakpoints is important, not only to the extent that it ensures good performance of the ensuing unit root tests, but also in its own right because researchers often seek to draw conclusions regarding these parameters.

5.4 Robustness and discussion

The results for alternative specifications of the common component are not reported but we describe them briefly. In agreement with theory, setting $\Pi \neq -I_m$ has no effect on the breakpoint estimator nor the unit root tests. In fact, the results are almost identical. Different values of λ_i have also no effect on test performance. The overall picture is therefore the same as the one given in Tables 1–3 for the case when $\Pi \neq -I_m$ and $\lambda_i \sim N(1, 1)$.

As we have seen, the outlier detection procedure seems quite robust to departures from a unit root, which is to be expected as the extreme value theory used for deriving the asymptotic distribution of the Wald test only requires stationary series. However, normality is crucial, and we have not yet seen the effects of non-normal errors. Table 4 therefore reports some results obtained when using alternative error distributions. For simplicity, in this case we set $\rho = 0$ with $\gamma = 0$, $r = 1$ and $\delta = 3$ kept fixed.

¹²An average break size of $\delta = 3$ is quite small in the sense that when viewed as an outlier it is only three standard deviations of the underlying innovations.

It is seen that heavy-tailed distributions make the breaks more difficult to detect with substantially lower correct selection frequencies in many cases, which is in agreement with the results obtained in most of the literature on outlier detection, see for example Perron and Rodríguez (2003). By contrast, the size of the unit root tests is basically unaffected, which is to be expected as the central limit theory used here does require any distributional assumptions, like normal.

As a final note, we simulated the Lee and Strazicich (2003) LM_τ test and its panel version, which was used by Im *et al.* (2005) in their PPP application. The results suggest that the size of these tests can be very unreliable with massive distortions in a majority of the cases considered. We will come back to this in the next section when we revisit the PPP hypothesis.

6 The PPP hypothesis revisited

The PPP hypothesis is the simple proposition that national price levels should tend to equalize when expressed in a common currency so that movements in the real exchange rate should only reflect stationary deviations from its long-run equilibrium level. Let p_{it} denote the local currency price index of country i expressed in log terms, and let s_{it} be the log dollar price of the currency of the same country. Also, let p_t^* denote the dollar price index. The log real exchange rate between country i and the United States can be written as

$$q_{it} = p_{it} + s_{it} - p_t^*.$$

Empirically, the stationarity of the real exchange rate has been relatively easy to evidence using data that span long periods of time. However, it has been considerably more difficult to find such evidence for the relatively short spans of data corresponding to the recent floating exchange rate period that followed the collapse of the Bretton Woods system in 1973. Consequently, studies such as Choi (2001, 2004) and Papell and Theodoridis (1998) try to remedy this absence of long-span data under the recent float by using unit root tests that are based on panel data. Yet, the results have been very mixed and far from convincing.

Im *et al.* (2005) argue that this weak empirical support may be due to the presence of structural breaks in the level of the equilibrium real exchange rate.¹³ If this is the case, conventional panel data tests based on the no-break assumption will suffer from low power,

¹³Similar explanations for why PPP does not seem to hold have been put forth by for example Gadea *et al.* (2004) and Sabatè *et al.* (2003).

which could explain the inability of earlier studies to reject the unit root null. To test this conjecture, the authors propose using their newly devised exogenous one-break panel unit root test.

However, since the breaks are unknown in this case, as they usually are, the authors suggest modifying their test along the lines of Lee and Strazicich (2003) using the Zivot and Andrews (1992) minimum procedure to estimate the breaks. Based on four different panels covering between six and 21 countries from April 1973 to December 1999, the authors are able to reject the unit root null at the 1% level, which interpreted as providing support in favor of the PPP hypothesis.

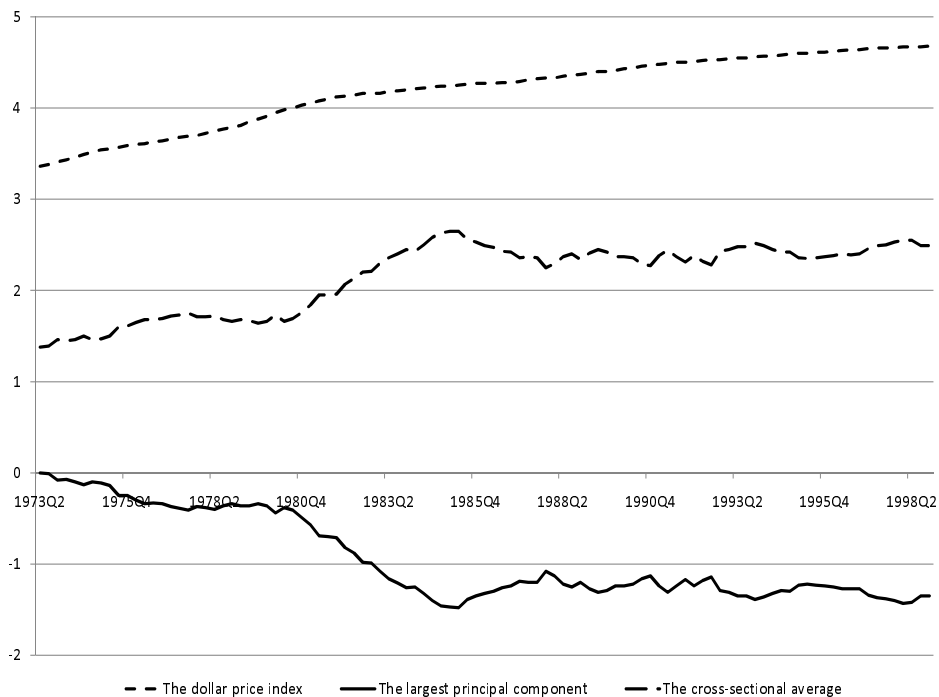
The results reported in this paper suggest that there is an alternative interpretation of these results. Namely, that they have been spuriously induced by the bias inherent in the Zivot and Andrews (1992) procedure when applied in this context. In this section, we therefore reevaluate the results of Im *et al.* (2005) using the new tests.

These tests have several advantages in comparison with conventional testing approaches. Firstly, structural breaks in both the level and slope of the PPP relationship can be taken into account. Secondly, by using panel data, we can improve upon the power relative to conventional univariate testing approaches. Thirdly, as first pointed out by O'Connell (1998), by permitting for common factors the cross-sectional dependence induced by the numeraire country is effectively eliminated. Finally, neither the constant nor the trend is restricted to zero. Usually there is only a constant, as a time trend is deemed inconsistent with PPP theory. The results obtained from this testing strategy have, however, been very mixed and far from convincing, leading Hegwood and Papell (1998) to consider so-called trend qualified PPP, which seems more appropriate in view of the trending behavior of most exchange rates. In addition, besides its better ability to explain the observed data, the allowance of the trend makes the model robust against possible Balassa-Samuelson effects.

Clearly, these allowances lead to a very relaxed interpretation of PPP under the stationary alternative. Thus, if there is only some limited truth to PPP, then our approach should in principle still be able to detect this. Moreover, as the null hypothesis is that of no PPP, an acceptance by the panel test not only suggests that this relaxed PPP version fails but also that it fails for each and every country in the panel. In other words, if the PPP hypothesis does not survive this very general test, then maybe it is time to consider revising the underlying theory?

Data sampled at quarterly frequency is obtained using the International Financial Statistics database of International Monetary Fund, which is the same source used by Im *et al.* (2005). As in that study, the countries are grouped into four panels, the Choi (2001) panel, the European monetary union panel, the European community panel and the OECD panel, which includes all 21 countries.¹⁴ The data starts in 1973Q2 but ends in 1998Q4 due to missing observations in 1999. Thus, in this panel, $N = 21$ and $T = 104$.

Figure 1: The estimated factors.



The tests are computed in the same ways as before, but since the common factors are now unknown in this section we will replace them by estimates. Three estimators are considered. The first is based on using the dollar price index, p_t^* , as an observed factor, which is very interesting in the sense that it provides an example of the scenario considered in Sections 2–5. The second estimator is based on using the cross-sectional average of the observed data, \bar{y}_t , while the third is based on using the principal components method of Bai and Ng (2004). The latter approach is more general in the sense that it allows us to extract more than one common factor from the data. But since the co-movement in the real exchange rate is mainly

¹⁴See Table 5 for a list of all the countries included in the sample.

driven by p_t^* , in this section we only extract one factor.¹⁵

The three factor estimates are plotted in Figure 1. Because the principal components estimate is only identified up to a scale factor, it is not directly comparable with the other two. However, it is seen that the principal components estimate is almost a perfect mirror image of the cross-sectional average, which in turn is quite close to the dollar price index. The three estimates therefore deliver roughly the same picture.

The results from applying the new tests to the PPP data are reported in Tables 5 and 6. There is strong evidence in favor of the unit root null. In fact, except for France when using the principal components method to estimate the common factor, the unit root hypothesis cannot be rejected at any conventional significance level regardless of whether we use the panel or the individual time series tests. We also tested for unit roots in the estimated factors using the Dickey–Fuller test. In agreement with the obvious non-stationarity of these series as seen in Figure 1, none of tests were able to reject the null of a unit root.

Of course, this not only casts doubts on the test results provided by Im *et al.* (2005), but also on the argument that the weak empirical support for PPP is be due to erroneously omitted breaks. In fact, as seen in Table 5, if we focus on the results based on the dollar price index as the common factor, the evidence of structural instability is weak. Also, many of the breaks detected for the other factor estimates can probably be attributed to the same underlying shift. Take for example Finland, where we estimate six breaks. But four of these, corresponding to observations 74, 77, 79 and 80, are very close, indicating the presence of a major break in the early 1990's, which seems roughly consistent with the formation of the European Monetary System when Finland, along with most European countries, abolished its capital controls.

Looking across all 21 countries we see that there is a preponderance of breaks between observation 11 and 40, corresponding to the 1975Q3–1982Q4 period. This seems very reasonable from an historical point of view with events such as oil price shocks, the formation of European Monetary System and, in particular, the rise and fall of the dollar. Specifically, while the breaks found in the late 1970's agree with the oil price shocks of that period, the breaks in the early 1980's are more likely to reflect the start of appreciation of the dollar.

In sum, our results lead us to the conclusion that PPP must be rejected for each country,

¹⁵Since the common factor model in (2) is basically the same as in Bai and Ng (2004), this suggests that we can use their proposal of first taking differences to achieve stationarity, and then to estimate the first-differenced factor using principal components, which can be accumulated and used in place of f_t .

and therefore for the panel as a whole, which is actually quite remarkable given that our model is so general.

7 Conclusions

This paper develops two Lagrange multiplier-based unit root tests that permit for multiple structural breaks in the level of the data. One is based on a single series, while the other is based on a panel of multiple time series. To estimate the breaks, a new procedure based on outlier detection is proposed. The new procedure has many distinctive and advantageous features. Firstly, it is computationally very simple and straightforward to implement. Secondly, neither the number nor the location of the breaks need be known. In fact, there may not be any breaks at all. Thirdly, the procedure is valid under both the unit root null and stationary alternative. Fourthly, the procedure is robust to both serial and cross-sectional dependence.

In essence, the new procedure allows researchers to move away from testing the unit root null against a specified number of breaks and towards model selection strategies that are less dependent on an prespecified number, which should be of considerable interest in applied work.

We derive the limiting distribution of the new tests and consider their finite-sample properties through a small simulation study. The results suggest that the asymptotic properties of the tests are borne out well in small samples, which together with the endogenous treatment of the breaks, leads us to the conclusion that the new tests should be a valuable addition to the existing menu of unit root tests. This is illustrated empirically using PPP, where it is shown that even if the presence of structural breaks and cross-section dependence is taken into account, the null of a non-stationary real exchange rate cannot be rejected.

References

- Ahn, S. K. (1993). Some Tests for Unit Roots in Autoregressive-Integrated-Moving Average Models with Deterministic Trends. *Biometrika* **80**, 855–868.
- Amsler, C., Lee, J. (1995). An LM Test for a Unit Root in the Presence of a Structural Break. *Econometric Theory* **11**, 359–368.
- Chareka, P., Matarise, F., Turner, R. (2006). A Test for Additive Outliers Applicable to Long-Memory Time Series. *Journal of Economic Dynamics and Control* **30**, 595–621.
- Chen, C., Liu, L.-M. (1993). Joint Estimation of Model Parameters and Outlier Effects in Time Series. *Journal of the American Statistical Association* **88**, 284–297.
- Choi, I. (2001). Unit Root Tests for Panel Data, *Journal of International Money and Finance* **20**, 249–272.
- Choi, C.-Y. (2004). Searching for Evidence of Long-Run PPP From a Post-Bretton Woods Panel: Separating the Wheat From the Chaff. *Journal of International Money and Finance* **23**, 1159–1186.
- Embrechts, P., Klüppelberg, C., Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Berlin, Springer.
- Gadea, M.-D., Montañés, A., Reyes, M. (2004). The European Union Currencies and the US Dollar: From Post-Bretton-Woods to the Euro. *Journal of International Money and Finance* **23**, 1109–1136.
- Hegwood, N. D., Papell, D. H. (1998). Quasi Purchasing Power Parity. *International Journal of Finance and Economics* **3**, 279–289.
- Im, K. S., Lee, J., Tieslau, M. (2005). Panel LM Unit Root Tests with Level Shifts. *Oxford Bulletin of Economics and Statistics* **67**, 393–419.
- Kapetanios, G. (2005). Unit-Root Testing Against the Alternative Hypothesis of up to m Structural Breaks. *Journal of Time Series Analysis* **26**, 123–133.
- Lee, J., Strazicich, M. (2003). Minimum LM Unit Root Test with Two Structural Breaks. *Review of Economics and Statistics* **85**, 1082–1089.

- O'Connell, P. G. J. (1998). The Overvaluation of Purchasing Power Parity. *Journal of International Economics* **44**, 1–19.
- Papell, D. H., Theodoridis, H. (1998). Increasing Evidence of Purchasing Power Parity Over the Current Float. *Journal of International Economics* **17**, 41–50.
- Perron, P. (1989). The Great Crash, the Oil Price Shock and the Unit Root Hypothesis. *Econometrica* **57**, 1361–1401.
- Perron, P., Rodríguez, G. (2003). Searching for Additive Outliers in Nonstationary Time Series. *Journal of Time Series Analysis* **24**, 193–220.
- Pesaran, M. H. (2007). A Simple Panel Unit Root Test in the Presence of Cross Section Dependence. *Journal of Applied Econometrics* **22**, 265–312.
- Sabatè, M., Gadea, M.-D., Serrano, J. M. (2003). PPP and Structural Breaks. The Peseta–Sterling Rate, 50 Years of a Floating Regime. *Journal of International Economics* **22**, 613–627.
- Schmidt, P., Phillips, P. C. B. (1992). LM Tests for a Unit Root in the Presence of Deterministic Trends. *Oxford Bulletin of Economics and Statistics* **54**, 257–287.
- Vogelsang, T. J. (1999). Two Simple Procedures for Testing for a Unit Root when there are Additive Outliers. *Journal of Time Series Analysis* **20**, 237–252.
- Zivot, E., Andrews, D. W. K. (1992). Further Evidence on the Great Crash, the Oil-Price Shock and the Unit-Root Hypothesis. *Journal of Business and Economics Statistics* **10**, 251–270.
- Westerlund, J., Edgerton, D. (2007). New Improved Tests for Cointegration with Structural Breaks. *Journal of Time Series Analysis* **28**, 188–224.

Table 1: Size for the unit root tests when $\rho = 0$.

δ	T	N	$r = 1$						$r = 2$					
			$\tau_{\rho_1}^1$	$\tau_{\rho_1}^2$	$\tau_{\rho_1}^3$	τ_{ρ}^1	τ_{ρ}^2	τ_{ρ}^3	$\tau_{\rho_1}^1$	$\tau_{\rho_1}^2$	$\tau_{\rho_1}^3$	τ_{ρ}^1	τ_{ρ}^2	τ_{ρ}^3
Independent errors, $\gamma = 0$														
3	100	5	6.1	5.6	6.3	8.9	8.5	9.2	6.6	6.5	6.8	10.4	11.4	9.9
	200	10	6.3	6.1	6.0	9.5	9.0	9.2	6.4	6.4	6.4	10.8	13.0	10.9
5	100	5	6.0	5.6	5.8	7.5	6.6	7.1	5.9	5.8	5.8	8.1	8.2	7.5
	200	10	5.6	5.5	5.7	7.8	7.6	8.1	5.6	5.6	5.8	8.2	9.8	8.6
3	100	5	6.5	5.6	6.3	8.8	8.6	9.2	7.0	7.0	6.8	9.9	15.3	9.9
	200	10	6.2	5.8	6.0	10.1	8.2	9.2	6.4	6.7	6.4	10.8	20.1	10.9
5	100	5	5.8	5.7	5.8	7.3	6.8	7.1	5.8	6.0	5.8	7.5	10.7	7.5
	200	10	5.7	5.5	5.7	8.2	7.4	8.1	5.8	5.9	5.8	8.4	13.7	8.6
Serially correlated errors, $\gamma = 0.5$														
3	100	5	5.9	5.3	6.3	7.9	6.8	8.4	5.5	4.9	6.3	7.7	6.2	9.1
	200	10	5.8	5.3	6.1	8.9	6.4	9.8	5.6	4.7	6.3	7.3	5.8	10.2
5	100	5	5.5	5.0	5.6	6.4	5.8	7.0	5.2	4.8	5.7	6.2	5.0	7.3
	200	10	5.6	5.2	5.8	7.3	6.1	7.7	5.4	4.9	5.8	6.1	5.2	8.0
3	100	5	6.0	4.4	6.3	8.0	5.2	8.4	5.8	4.0	6.3	8.1	5.0	9.1
	200	10	6.1	4.4	6.1	8.4	3.6	9.8	6.1	3.9	6.3	8.3	4.0	10.2
5	100	5	5.6	4.5	5.6	6.8	4.6	7.0	5.5	4.0	5.7	6.8	4.0	7.3
	200	10	5.7	4.7	5.8	7.4	4.7	7.7	5.5	4.1	5.8	7.2	3.5	8.0

Notes: The value γ refers to first-order autoregressive parameter in the errors, δ refers to the average break size, and r refers to the number of breaks. The superscripts 1, 2 and 3 in τ_{ρ_i} and τ_{ρ} refer to the tests based on estimated, omitted and known breaks, respectively.

Table 2: Size-adjusted power for the unit root tests when $\rho = -0.1$.

δ	T	N	$r = 1$						$r = 2$					
			$\tau_{\rho_i}^1$	$\tau_{\rho_i}^2$	$\tau_{\rho_i}^3$	τ_{ρ}^1	τ_{ρ}^2	τ_{ρ}^3	$\tau_{\rho_i}^1$	$\tau_{\rho_i}^2$	$\tau_{\rho_i}^3$	τ_{ρ}^1	τ_{ρ}^2	τ_{ρ}^3
Independent errors, $\gamma = 0$														
3	100	5	7.9	7.2	9.3	62.9	58.1	70.8	7.5	7.5	8.1	59.6	61.8	68.3
		10	4.2	4.0	4.4	90.1	85.0	94.6	4.0	4.0	4.5	87.8	89.3	93.8
	200	5	31.0	28.5	37.0	100.0	99.8	100.0	29.8	29.6	36.6	99.8	99.8	100.0
		10	21.9	19.5	25.8	100.0	100.0	100.0	22.1	21.9	25.6	100.0	100.0	100.0
5	100	5	8.1	4.7	9.3	63.3	39.5	70.8	7.5	6.7	8.1	63.5	52.1	68.3
		10	4.5	3.0	4.4	92.1	67.7	94.6	4.0	3.3	4.5	89.0	80.9	93.8
	200	5	33.4	17.7	37.0	100.0	98.3	100.0	31.0	22.9	36.6	99.8	98.9	100.0
		10	24.6	12.4	25.8	100.0	100.0	100.0	23.2	15.0	25.6	100.0	100.0	100.0
Serially correlated errors, $\gamma = 0.5$														
3	100	5	7.1	7.0	7.5	61.5	57.3	64.7	6.8	6.9	6.8	61.3	61.6	63.6
		10	4.2	4.2	4.6	88.8	87.8	89.2	4.5	4.5	4.5	87.8	89.2	88.6
	200	5	31.8	30.9	34.1	99.7	99.7	99.8	31.4	32.2	33.9	99.7	99.8	99.8
		10	21.4	20.5	21.1	100.0	100.0	100.0	21.6	21.1	21.5	100.0	100.0	100.0
5	100	5	7.1	6.8	7.5	61.4	52.3	64.7	6.7	7.0	6.8	59.6	59.4	63.6
		10	4.5	4.0	4.6	88.8	83.5	89.2	4.1	4.1	4.5	87.5	88.3	88.6
	200	5	32.6	26.3	34.1	99.7	99.5	99.8	31.9	28.8	33.9	99.8	99.7	99.8
		10	21.1	19.1	21.1	100.0	100.0	100.0	20.1	21.2	21.5	100.0	100.0	100.0

Notes: See Table 1.

Table 3: Breakpoint estimation accuracy.

δ	T	N	$\rho = 0$				$\rho = -0.1$			
			$r = 1$		$r = 2$		$r = 1$		$r = 2$	
			\hat{r}	\hat{B}	\hat{r}	\hat{B}	\hat{r}	\hat{B}	\hat{r}	\hat{B}
Independent errors, $\gamma = 0$										
3	100	5	37.1	33.2	21.7	19.7	35.9	31.9	20.0	18.2
		10	38.1	34.1	21.9	20.1	36.5	32.4	20.4	18.6
	200	5	36.3	32.3	21.0	19.3	34.7	30.5	19.3	17.6
		10	35.4	31.5	20.3	18.6	33.8	29.8	18.5	17.0
5	100	5	74.8	74.0	65.0	64.1	73.2	72.3	62.7	61.8
		10	75.4	74.4	65.4	64.4	73.5	72.5	63.2	62.1
	200	5	76.5	75.5	66.8	65.9	74.6	73.5	64.7	63.5
		10	75.6	74.7	65.7	64.7	73.7	72.7	63.1	61.9
Serially correlated errors, $\gamma = 0.5$										
3	100	5	36.4	32.0	20.3	18.2	35.7	31.4	19.4	17.5
		10	37.3	32.8	20.8	18.7	36.4	32.0	19.7	17.9
	200	5	35.4	30.8	19.8	17.5	34.2	29.7	18.8	16.9
		10	34.5	30.2	19.3	17.1	33.4	29.1	18.2	16.3
5	100	5	71.2	69.9	61.8	60.3	71.9	70.6	61.1	59.6
		10	71.4	70.0	61.9	60.4	71.9	70.5	60.9	59.5
	200	5	69.8	68.3	60.7	59.0	71.5	70.1	61.2	59.8
		10	69.1	67.7	59.6	57.7	70.7	69.5	60.2	58.6

Notes: \hat{r} and \hat{B} refer to the estimated number of breaks and their locations, respectively. The numbers reported in the table are the percentage of times when these parameters are correctly selected. See Table 1 for an explanation of the rest of the features.

Table 4: Simulation results for different error distributions.

T	N	Unit roots						Breaks	
		$\tau_{\rho_i}^1$	$\tau_{\rho_i}^2$	$\tau_{\rho_i}^3$	τ_{ρ}^1	τ_{ρ}^2	τ_{ρ}^3	\hat{r}	\hat{B}
$\epsilon_{it} \sim t_2$									
100	5	7.5	5.2	5.3	10.9	5.8	6.5	17.9	0.6
	10	7.0	5.3	5.2	11.4	5.8	5.9	17.5	0.5
200	5	6.6	4.7	4.8	9.1	5.3	5.8	3.0	0.0
	10	6.4	4.2	4.4	8.7	4.5	5.0	3.2	0.0
$\epsilon_{it} \sim t_{10}$									
100	5	6.8	6.5	6.6	8.8	8.3	8.6	33.4	17.7
	10	6.5	6.3	6.5	8.9	8.0	10.0	34.2	18.4
200	5	5.7	5.5	5.9	8.4	7.3	7.6	35.7	14.2
	10	5.7	5.6	5.9	7.5	7.6	7.8	34.8	13.5
$\epsilon_{it} \sim U(-1,1)$									
100	5	6.2	5.7	6.1	8.4	7.0	8.4	78.0	78.0
	10	6.3	5.9	6.2	10.5	8.3	10.3	78.0	78.0
200	5	5.4	5.1	5.5	7.0	6.2	6.9	77.0	77.0
	10	5.9	5.7	5.8	8.8	6.8	8.9	76.2	76.2
$\epsilon_{it} \sim U(-2,2)$									
100	5	5.9	5.6	6.1	8.1	7.7	8.4	24.5	24.5
	10	6.2	6.0	6.2	9.8	8.9	10.3	24.6	24.6
200	5	5.4	5.2	5.5	6.5	6.5	6.9	21.6	21.6
	10	5.7	5.6	5.8	8.3	7.3	8.9	21.4	21.4

Notes: See Tables 1 and 3.

Table 5: Country-specific results.

	Unit root tests			Estimated breakpoints		
	$\tau_{\rho_i}^1$	$\tau_{\rho_i}^2$	$\tau_{\rho_i}^3$	$\hat{\beta}^1$	$\hat{\beta}^2$	$\hat{\beta}^3$
Australia	-1.69	-2.17	-2.17	14	14	14
Austria	-2.66	-1.24	-1.23	0	0	0
Belgium	-1.76	-1.43	-1.55	0	0	0
Canada	-1.20	-1.61	-1.65	0	0	0
Denmark	-1.62	-1.66	-2.05	0	0	0
Finland	-1.80	-1.61	-1.72	0	16, 38, 74, 77, 79, 80	16, 38, 74, 77, 79, 80
France	-1.55	-3.12*	-2.70	0	13, 37	0
Germany	-2.58	-1.38	-1.35	0	0	0
Greece	-1.04	-0.86	-0.86	78	28, 39, 50	28, 39, 50
Iceland	-1.25	-1.12	-1.12	37	37	37
Ireland	-1.40	-1.56	-1.54	0	11	11
Italy	-1.60	-2.43	-2.28	11	11, 77, 80, 87	11, 77, 80, 87
Japan	-1.63	-1.57	-1.64	0	89	89
Luxembourg	-1.80	-1.35	-1.45	0	0	0
Netherlands	-1.91	-1.00	-1.11	0	37	0
Norway	-1.46	-1.60	-1.71	78	0	0
Portugal	-1.04	-1.29	-1.28	15	11, 15, 20, 27, 36, 37	11, 15, 20, 27, 36, 37, 40
Spain	-1.58	-1.88	-1.80	17	11, 17, 37, 38	11, 17, 38
Sweden	-1.42	-2.51	-2.87	78	38, 78	38, 78
Switzerland	-2.67	-1.11	-1.10	0	22	22
United Kingdom	-1.67	-1.61	-1.72	0	0	0

Notes: The superscript 1 indicates that the results have been obtained by using the dollar price index as the common factor, 2 indicates that the factor has been estimated using the principal components method, 3 indicates that the factor has been estimated using the cross-sectional averages of the data, and * indicates significance at the 5% level. See Table 1 and 3 for an explanation of the rest of the features.

Table 6: Panel unit root tests.

Panel	τ_{ρ}^1	p -value	τ_{ρ}^2	p -value	τ_{ρ}^3	p -value
Choi (2001)	0.83	0.80	2.07	0.98	1.85	0.97
European monetary union	1.47	0.93	2.87	1.00	2.83	1.00
European community	1.70	0.96	2.34	0.99	2.58	1.00
OECD	2.31	0.99	2.78	1.00	2.48	0.99

Notes: See Table 5.