Simple Unit Root Tests with Multiple Breaks*

Joakim Westerlund†

December 14, 2006

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 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; Cross-Sectional Dependence; 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 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 nonstationarity of the real exchange

*The author would like to thank Anindya Banerjee, David Edgerton and seminar participants at Lund University 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 Depertment 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. The usual disclaimer applies.

†Department of Economics, Lund University, P. O. Box 7082, S-220 07 Lund, Sweden. Telephone: +46 46 222 8670; Fax: +46 46 222 4118; E-mail address: joakim.westerlund@nek.lu.se.

rate has been frequently documented, see for example Choi (2004) for a good review of the recent literature.

An important feature common to most unit root studies 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.

In a recent article, Amsler and Lee (1995) remedy this critic by developing a new test based on the Lagrange multiplier (LM) 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 break. The authors show that the asymptotic distribution of the test does not depend on the usual nuisance parameter representing the location of the break. This unusual feature is of course very convenient as it implies 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 Zivot and Andrews (1992) break estimation procedure. Specifically, a two-break test is suggested where the breakpoints are estimated at the minimum of the test statistics from across all allowable break dates. However, fixing the number of breaks to two is as arbitrary as one, and there is a need to permit not only the locations but also the number of breaks to be unknown. Moreover, for reasons to be explained later, the minimum LM test is expected to suffer from distortions in small samples.

In this paper, we build on the results of Amsler and Lee (1995) and generalize their test 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. First, it is simple and computationally less intensive than other procedures such as that of Zivot and Andrews (1992). Second, it valid under both the unit root null and stationary alternative, which makes the outcome of the test easy to interpret.

Moreover, given the potential loss of power from ignoring 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 test 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. Section 2 presents the new time series test and its limiting distribution under the null of a unit root with exogenous breaks. The test is then extended first in Section 3 to accommodate unknown breaks, and again in Section 4 to handle the case with panel data. Section 5 is concerned with the Monte Carlo study whereas Section 6 contains the empirical application. Section 6 concludes.

2 A unit root test with multiple breaks

In this section, we generalize the unit root test of Amsler and Lee (1995) to allow for multiple breaks. Towards this end, we will assume that the time series variable y_t, indexed $t = 1, ..., T$, can be described as

$$y_t = \alpha + \tau t + \sum_{j=1}^{K} \delta_j D_t(T_j^b) + z_t, \tag{1}$$

$$\Delta z_t = \phi z_{t-1} + \sum_{j=1}^{p} \gamma_j \Delta z_{t-j} + e_t. \tag{2}$$

The dummy variables $D_t(T_j^b)$ represent the structural breaks. In particular, if we let T_j^b denote the location of break number $j = 1, ..., K$, then $D_t(T_j^b) = 1$ if $t > T_j^b$ and zero otherwise. Thus, in this model, α represents the level of the series before any break takes place and δ_j represents the change in the level at the time of break j. The disturbance e_t is assumed to be mean zero and serially uncorrelated.

In this simple setup, the problem of testing the null hypothesis of a unit root is equivalent to testing $H_0 : \phi = 0$ versus $H_1 : \phi < 0$, which, as shown by Amsler and Lee (1995), can be implemented using the following regression

$$\Delta \widehat{S}_t = \text{constant} + \phi \widehat{S}_{t-1} + \sum_{j=1}^{p} \gamma_j \Delta \widehat{S}_{t-j} + \text{error}. \tag{3}$$

The variable $\widehat{S}_t$ is defined as

$$\widehat{S}_t \;=\; y_t - \widehat{\alpha} - \widehat{\tau} t - \sum_{j=1}^{K} \widehat{\delta}_j D_t(T_j^b),$$

where $\widehat{\alpha} = y_1 - \widehat{\tau} - \sum_{j=1}^{K} \widehat{\delta}_j D_1(T_j^b)$ is the restricted maximum likelihood estimate of α under the null hypothesis. The corresponding estimates $\widehat{\tau}$ and $\widehat{\delta}$ of τ and δ are obtained by running least squares on (1) in first differences. That is, $\widehat{\tau}$ and $\widehat{\delta}$ are obtained from the following least squares regression

$$\Delta y_t \;=\; \widehat{\tau} + \sum_{j=1}^{K} \widehat{\delta}_j \Delta D_t(T_j^b) + \text{error}. \tag{4}$$

As pointed out by Schmidt and Phillips (1992), a natural way to test the hypothesis of H_0 versus H_1 is to use the conventional t-statistic for testing the zero restriction on ϕ in (3).[1] Let us denote this statistic by τ_ϕ. To be able to obtain the limit distribution of τ_ϕ, we need to ensure that the breaks are asymptotically distinct. For this reason, it is assumed that T_j^b is set as a fixed fraction $\lambda_j \in [0,1]$ of T such that $T_j^b = \lambda_j T$ and $\lambda_j > \lambda_{j-1}$ for all j. The limit is taken as $T \to \infty$ in a sequence that ensures an integer value of T_j. For now, K and λ_j will be treated as known suggesting an completely exogenous break structure.

Under these conditions, it is possible to show that

$$\tau_\phi \;\Rightarrow\; -\frac{1}{2}\left(\int_0^1 U(r)^2 \mathrm{d}r\right)^{-1/2} \quad \text{as} \quad T \to \infty, \tag{5}$$

where $W(r)$ is a standard Brownian motion on the unit interval $r \in [0,1]$, $V(r) = W(r) - rW(1)$ is a Brownian bridge and $U(r) = V(r) - \int_0^1 V_i(s)\mathrm{d}s$ is a demeaned Brownian bridge. The proof of (5) proceeds in the same fashion as in Ahn (1993) and Amsler and Lee (1995), and is therefore omitted

There are several things about (5) that are worthy of further discussion. Firstly, since $U(r)$ is independent of K and λ_j, the distribution of the test is invariant with respect to both the number of breaks as well as their locations under the null hypothesis. This is of cause 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 where no breaks at all. If τ_ϕ is smaller than the appropriate critical value, the unit root null is rejected.

Secondly, since the breaks are allowed under both the null and alternative hypotheses, there is no confusion about the interpretation of the test outcome.

[1] Another advantage with choosing H_0 and H_1 in this way is that it circumvents the well-known problem of size distortion when testing the null of stationarity in the presence of serial correlation, see for example Craner and Kilian (2001).

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 nonstationary with breaks.

Similarly, 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 nonstationary real exchange rate with breaks.

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

In this section, we relax the assumption of known breaks. In so doing, we propose a very simple detection procedure that can be employed to estimate the breaks from the data.

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 K breaks is clearly computationally extremely demanding and practically infeasible for $K > 3$.[2] Another drawback of this approach is that K must be known. Thus, it is not possible to test whether the number of breaks in fact is equal to K or not.

Moreover, with an invariant test such as ours, this method may well result in an overrejection of the unit root null in small samples. As an example, consider the Lee and Strazicich (2003) proposal, which, in their model A, amounts to

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

applying the Zivot and Andrews (1992) procedure to the two-break τ_ϕ test. The authors argue that because the test is invariant, taking the minimum should not affect asymptotic distribution of the test. The critical values should therefore be the same as for the exogenous test of Schmidt and Phillips (1992).

However, this is not entirely true. The problem is that by taking the minimum, the authors are essentially taking an order statistic and treating it as an ordinary test statistic, which is likely to result in size distortions. The intuition is that because the same critical values are used, the minimum statistic will tend to be too large in absolute value, thus making the test biased towards rejecting the unit root null. Moreover, since the test does not depend on the break, the precision of the estimated breakpoints is likely to be very poor.

A more feasible approach that does not require the knowledge of K 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 from (3). This approach is taken by Kapetanios (2005), who proposes a vary efficient grid search scheme that requires only $O(T)$ least squares operations for any K. 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.[3]

An even simpler approach, that is perhaps more natural in our case, is to treat the estimation problem as one of outlier detection, and to estimate the breaks from the first differenced regression in (4), which is always stationary, rather than from (3). The standard approach to do this is to estimate (4) and 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 can be described as follows. With a single outlier at time T^b of magnitude δ, the model in differences can be written as

$$\Delta y_t \;=\; \tau + \delta \Delta D_t(T^b) + \Delta z_t, \tag{6}$$

where $\Delta D_t(T^b) = 1$ if $t = T^b$ and zero otherwise, and the disturbance Δz_t is defined in (2). If we let N denote the ratio $(T-2)/(T-1)$, then the least squares estimate of δ in the above regression can be calculated as

$$\widehat{\delta}(T^b) \;=\; \frac{1}{N}\left(\Delta y_{T^b} - \frac{y_T - y_1}{T-1}\right).$$

Now, consider testing the null hypothesis of no outlier. Under this null, the term within parenthesis reduces to $\Delta z_{T^b} - (z_T - z_1)/(T-1)$. This implies that the variance of $\widehat{\delta}(T^b)$ can be formulated as $E(\Delta z_t^2)/N^2$, which can be readily estimated using $\widehat{\sigma}^2/N^2$, where $\widehat{\sigma}^2$ is the estimated regression variance from (6).

[3] Sabatè *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.

By using these results, we have that the t-statistic for testing $\delta = 0$ can be computed as $\tau_\delta(T^b) = \widehat{\delta}(T^b)/\widehat{\sigma}$.

Our test statistic for the null of no outlier is given by

$$\tau_\delta \;=\; \sup_{T^b} |\, \tau_\delta(T^b)\,|.$$

The limit distribution of τ_δ is invariant to nuisance parameters such as serial correlation and heteroskedasticity but it depends on the specific distribution of e_t, which is exactly the same problem as in testing for outliers in stationary time series. The standard practice in the literature is quite arbitrary and consists of rejecting the no outlier null if τ_δ is greater than some critical value between 3 and 4. As a response to this arbitrariness, Perron and Rodrígues (2003) simulates critical values under the unit root null, which is shown to work well even in very small samples.

By using on these critical values, our one-outlier detection procedure can be implemented as follows. First, compute τ_δ for the entire series and compare the it to the appropriate critical value from Table IV in Perron and Rodrígues (2003). If t_δ exceeds this value, then an outlier is detected at date

$$\widehat{T}^b \;=\; \arg\max_{T^b} |\, \tau_\delta(T^b)\,|.$$

To detect multiple outliers, we can follow a strategy similar to that suggested by Vogelsang (1999), in which observations labelled as outliers are sequentially dropped from (6). Once the outlier observation has been removed, (6) is again estimated and tested for an outlier. This continues until the test fails to reject. Moreover, because we are dealing with differenced series, which are stationary, τ_δ is asymptotically independent at each step of the iterations. Thus, the same critical values apply at each step.

Another advantage with this procedure is that K may be unknown. In fact, K may be zero. If no outlier is detected, the unit root testing is carried out as described in Section 2 with $\widehat{S}_t$ based on no dummy variables, whereas, if there are outliers, then $\widehat{S}_t$ is augmented with one dummy variable $D_t(\widehat{T}^b_j)$ for each estimated break date $\widehat{T}^b_j$. This is of course very different from the Zivot and Andrews (1992) procedure, in which a nonzero value of K has to be stipulated beforehand.

Yet another advantage with using τ_δ 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 nonstationary.

4 A panel data test

The results of the previous sections can be easily generalized to handle the case with panel data, in which each series is observed over a cross-section of similar units such as individuals, industries or countries. For this reason, the index $i = 1, ..., N$ will be used to denote the cross-sectional units, while t again denotes time. For simplicity, the panel is assumed to be balanced so that each unit has the same number of time series observations, T.

The panel model that we consider is given by

$$y_{it} = \alpha_i + \tau_i t + \sum_{j=1}^{K_i} \delta_{ij} D_t(T_{ij}^b) + z_{it}, \quad (7)$$

$$\Delta z_{it} = \phi_i z_{it-1} + \sum_{j=1}^{p_i} \gamma_{ij} \Delta z_{it-j} + e_{it}. \quad (8)$$

Thus, we basically assume here that each of the cross-sectional units evolves according to (1) and (2). Note that no restrictions are placed on p_i, the order of the serial correlation, or K_i, number of breaks, which are both permitted to be completely heterogeneous. Similarly, there are no restrictions on the heterogeneity of the locations of the breaks.

Our panel model is thus very flexible. In fact, the only restriction is that the disturbance e_{it} is uncorrelated across i, which is a fairly common assumption when dealing with panel data. Of course, in applied work, this assumption can be quite restrictive, especially when testing PPP, and Section 6 therefore suggests a simple bootstrap procedure for dealing with the case of cross-sectionally correlated data. However, for the present, we retain the no correlation assumption and focus on the structural break issue.

The panel test that we propose generalizes the exogenous one-break test studied by Im *et al.* (2005). Our interest lies in testing the null hypothesis that all units are nonstationary versus the alternative that there is at least one unit that is stationary, which can be expressed as $H_0 : \phi_i = 0$ for all i versus $H_1 : \phi_i < 0$ 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.[4]

For testing this hypothesis, we propose using the cross-sectional average of the individual τ_ϕ statistics for each cross-section, denoted $\tau_{\phi N}$. The limit distribution of this test is readily deduced by using the independence of the cross-section, from which it follows that, as $T \to \infty$ followed by $N \to \infty$

$$\frac{\sqrt{N}(\tau_{\phi N} - E(\tau_\phi))}{\sqrt{\mathrm{var}(\tau_\phi)}} \Rightarrow N(0,1),$$

where $E(\tau_\phi)$ and $\mathrm{var}(\tau_\phi)$ are the expected value and variance, respectively, of the limiting expression in (5). Numerical values of $E(\tau_\phi)$ and $\mathrm{var}(\tau_\phi)$ can be

[4] As 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.

obtained using simulations. 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 (5) and then to compute the moments. The simulated expectation and variance based on this method are -1.96901 and 0.32276, respectively.[5]

The test is implemented in two steps. The first step is to compute the individual τ_ϕ tests using the outlier detection procedure to estimate the number of breaks and their locations for each cross-section. In the second step, the average of the individual tests is taken to obtain $\tau_{\phi N}$, which is then standardized by $E(\tau_\phi)$ and $\mathrm{var}(\tau_\phi)$. If the resulting standardized test is smaller than the appropriate left tail critical value from the normal distribution, the null is rejected.

5 Monte Carlo simulations

In this section, we briefly investigate some of the small-sample properties of the new tests by means of simulations, which are organized with one experiment for the time series test and one for the panel test. All results are based on $2,000$ replications, where the first 50 time series observations in each replication is discarded to avoid possible initial value effect. The significance level is set to 5% throughout, and all powers are adjusted for size. Computational work is carried out in GAUSS.

5.1 Time series testing

The data is generated according to

$$\begin{aligned} y_t &= \sum_{j=1}^{K} \delta D_t(T_j^b) + z_t, & (9) \\ \Delta z_t &= \phi z_{t-1} + \gamma \Delta z_{t-1} + e_t, & (10) \end{aligned}$$

where $e_t \sim N(0,1)$. For the structural breaks, we have two cases depending on whether K is assumed to be known or not. In the first, $K = 1$ is known but the location is unknown, while in the second, both the number of breaks and their locations are unknown. In the latter, two values of K are considered, 1 and 2. When K is known, the break location λ_1 can be either 0.3 and 0.7, whereas when it is unknown, $\lambda_1 = 0.3$ and $\lambda_2 = 0.7$ are kept fixed. Results are reported for breaks of magnitude 2 and 5. For the size simulations, $\phi = 0$, and for the power simulations, $\phi < 0$.

[5]Im *et al.* (2005) uses 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.

In constructing the unit root test, 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 Table IV in Perron and Rodrígues (2003) was used. Also, for τ_ϕ, the critical value -3.02 was taken from Table 1A in Schmidt and Phillips (1992). To eliminate the endpoints, 10% of the sample was truncated at both beginning and end.

We begin by considering the performance of the test when K is known. In this case, we are interested in comparing τ_ϕ with the LM_τ test of Lee and Strazicich (2003), which also assumes a known number of breaks but estimates their locations at the minimum of the individual test statistics. Results are also reported for the τ_ϕ test based on the true breakpoint and for the Schmidt and Phillips (1992) $\widetilde{\tau}$ test with no break. To be able to discriminate between the τ_ϕ tests, we use $\widehat{\tau}_\phi$ to denote the test based on an estimated break. The results are presented in Table 1.

Looking first the size results, we see that $\widetilde{\tau}$ generally performs best, which is not unexpected given that the asymptotic null distribution of the test is independent of the break. Hence, even though there is a break in the data generating process, $\widetilde{\tau}$ is actually expected to perform well here. The $\widehat{\tau}_\phi$ test, which accounts for the break, is almost as accurate as $\widetilde{\tau}$, and performs only slightly worse. None of the tests appear to be affected much by the location of the break. This is also consistent with theory. As expected, LM_τ is generally quite oversized, even in samples with T as large as 200.

The power results generally coincide with what might be expected from the asymptotic theory. Firstly, the power increases with T, which is presumably a reflection of the consistency of the tests. Secondly, the power increases as ϕ departs form its hypothesized value under the null. Thirdly, the no-break $\widetilde{\tau}$ test is generally the least powerful, thus corroborating the result that erroneous omission of breaks should affect the test by lowering its power. Fourthly, $\widehat{\tau}_\phi$ is almost as powerful as τ_ϕ, which suggests that the loss of power involved in estimating the break rather than treating it as known is very small.

Next, we consider the results for the case with multiple breaks, which are provided in Table 2. In this case, we drop LM_τ and $\widetilde{\tau}$, and concentrate on $\widehat{\tau}_\phi$ and τ_ϕ. It is seen that both tests generally perform well under the null with only small distortions. The power is generally good and it quickly increases as T grows. As in the one-break case, τ_ϕ is most powerful.

As for the accuracy of the estimated breaks presented in Table 3, we make the following two observations. First, the accuracy is not affected much by the value of ϕ, which is consistent with our outlier detection procedure being valid under both the unit root null and stationary alternative. Second, as expected the accuracy depends on the magnitude of the break, and can be sometimes be very poor when $\delta = 2$. However, it should be noted that a break of this magnitude is quite small in the sense that, when viewed as an outlier, it is only

2 standard deviations of the underlying innovations. With reasonably sized breaks, accuracy is generally very good.

5.2 Panel testing

In this case, we use (9) and (10) to generate data for N cross-sectional units. For simplicity, all parameters are assumed to be the same for all units. Moreover, since our panel test is nothing but an average of individual τ_ϕ statistics, many of the conclusions drawn in the previous section applies here too. Therefore, since the effects of serial correlation were very small, γ is set to zero.

The individual τ_ϕ statistics used in obtaining $\tau_{\phi N}$ are computed exactly as described in Section 5.1. The standardized test is computed using the simulated moments from Section 4 and compared to the 5% critical value -1.645 obtained from the normal distribution.

The results contained in Table 4 can be summarized as follows. Firstly, size accuracy is generally quite high with only small deviations from the nominal 5% level. In Section 5.1, for small values of T, we observed a slight tendency for τ_ϕ to become oversized. With panel data, such distortions not only remain but have a tendency of accumulating and to become rather serious as N grows. Although there are no large distortions, this effect is nonetheless clearly visible in Table 4.

Secondly, the power is very good, and it increases quickly as both N and T grow. In particular, it is interesting to compare the results of Table 4 with those reported in Section 5.1. The difference is striking. To take one example, with two unknown breaks of magnitude 2, $\gamma = 0$ and $T = 100$, the power of τ_ϕ is about 20%. The power of $\tau_{\phi N}$ in this case with 10 cross-sections is 92%, an increase by almost a factor of five. Thus, there are potentially large power gains to be made by exploring the cross-sectional dimension.

As a final note, we simulated the panel version of the Lee and Strazicich (2003) LM_τ test, which was used by Im *et al.* (2005) in their PPP application. The results suggest that the size of this test can be very unreliable with rejection frequencies that can be up to 10 times the nominal level. 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 deviation 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), Choi (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.[6] If this is the case, then conventional panel data tests based on no break will suffer from low power, which could explain the inability of earlier studies to reject the unit root null. To test this conjuncture, 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 overwhelming support in favor of the PPP hypothesis.

The results provided 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. Therefore, in this section, we reevaluate the results of Im *et al.* (2005) using the new tests.

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 data is grouped into four panels, the Choi (2001) panel, the European monetary union panel, the European community panel and the OECD panel. The data starts in April 1973 but ends in December 1998 due to missing observations in 1999.

The results presented in Tables 5 and 6 suggest that the unit root null cannot be rejected at any conventional significance level based on the asymptotic p-values regardless of whether we use the panel or time series test. 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 might be due to

[6] Similar 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).

erroneously omitted breaks. In fact, as seen in Table 5, in only seven out of the 21 countries do we find evidence of structural instability.

As mentioned in Section 4, our panel test relies on assuming independence, or at least zero correlation, among the cross-sectional units, which is unlikely to hold in the current application. Therefore, to be able to test the robustness of our findings, we employ the bootstrap approach, which makes inference possible even under very general forms of cross-sectional dependence. The particular bootstrap opted for in this section resembles that used by Chang (2004) and proceeds as follows.

The first step is to fit the following least squares regression

$$y_{it} \;=\; \widehat{\alpha}_i + \widehat{\tau}_i t + \sum_{j=1}^{K_i} \widehat{\delta}_{ij} D_t(\widehat{T}_{ij}^b) + \widehat{z}_{it}, \qquad (11)$$

where $\widehat{T}_{ij}^b$ is obtained as described in Section 3. Having obtained the residual $\widehat{z}_{it}$, we then estimate the autoregression

$$\Delta\widehat{z}_{it} \;=\; \sum_{j=1}^{p_i} \widehat{\gamma}_{ij} \Delta\widehat{z}_{it-j} + \widehat{e}_{it}. \qquad (12)$$

This gives us the residual vector $\widehat{e}_t = (\widehat{e}_{1t}, ..., \widehat{e}_{Nt})'$, which forms the basis for our bootstrap. It is important to note that $\widehat{e}_t$ is obtained from equation (12) with the null hypothesis imposed. Otherwise, if the null is not imposed, this will render the subsequent bootstrap tests inconsistent.

We then generate bootstrap innovations e_t^* by sampling with replacement the centered residual vector

$$\widetilde{e}_t \;=\; \widehat{e}_t - \frac{1}{T}\sum_{j=1}^{T} \widehat{e}_j.$$

Note that by resampling $\widetilde{e}_t$ rather than $\widetilde{e}_{it}$, we are in effect preserving the cross-sectional correlation structure of $\widehat{e}_{it}$.

The next step is to generate the bootstrap sample y_{it}^*. This is accomplished by first constructing the bootstrap version of the error Δz_{it} as

$$\Delta z_{it}^* \;=\; \sum_{j=1}^{p_i} \widehat{\gamma}_{ij} \Delta z_{it-j}^* + e_{it}^*.$$

For initial values, we may use the first p_i observations of $\Delta\widehat{z}_{it}$. However, as pointed out by Chang (2004), this does not guarantee that Δz_{it}^* is stationary. An alternative approach, which makes Δz_{it}^* more stationary, is to generate a larger number, $T + n$ say, of Δz_{it}^* and then discard the first n values. Because this makes the initiation unimportant, we may simply use zeros to start up the recursion.

Next, we generate z_{it}^* recursively from Δz_{it}^* as

$$z_{it}^* = z_{i0}^* + \sum_{j=1}^{t} \Delta z_{ij}^*,$$

which again requites initiation through z_{i0}^*. The value zero will do. Finally, the bootstrap sample y_{it}^* is obtained as

$$y_{it}^* = \widehat{\alpha}_i + \widehat{\tau}_i t + \sum_{j=1}^{K_i} \widehat{\delta}_{ij} D_t(\widehat{T}_{ij}^b) + z_{it}^*,$$

Having obtained the bootstrap sample y_{it}^*, we then obtain the bootstrapped $\tau_{\phi N}$ statistic, which is constructed in the same way as its sample counterpart. The bootstrap test is implemented using $1,000$ bootstrap replications, which are used to compute p-values for $\tau_{\phi N}$ under the unit root null. The results presented in Table 6 indicate that our earlier conclusions based on the asymptotic normal distribution are not altered when the cross-sectional correlations are taken into account. Thus, we do not find any evidence of PPP.

7 Conclusions

This paper develops two LM based unit root tests that permit for multiple structural breaks in the level of the observed time series. One is based on a single series while the other is based on a panel of 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.

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 unit root tests and consider their small-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 nonstationary real exchange rate cannot be rejected.

Table 1: Size and power for an known number of breaks.

				$\phi = 0$				$\phi = -0.1$				$\phi = -0.2$			
γ	λ	δ	T	$\widehat{\tau}_\phi$	LM_τ	$\widetilde{\tau}$	τ_ϕ	$\widehat{\tau}_\phi$	LM_τ	$\widetilde{\tau}$	τ_ϕ	$\widehat{\tau}_\phi$	LM_τ	$\widetilde{\tau}$	τ_ϕ
0	0.3	2	50	7.8	18.1	6.8	9.5	7.4	8.3	8.4	9.3	14.1	16.2	11.7	19.1
			100	6.6	13.0	6.1	6.5	18.1	19.2	19.3	22.7	40.7	48.3	42.6	61.4
			200	5.3	11.4	4.9	5.3	60.8	65.2	62.0	70.0	94.6	97.7	96.2	98.6
		5	50	8.5	19.4	6.3	8.7	7.8	6.2	6.9	8.2	16.5	14.1	9.8	18.5
			100	5.7	14.3	6.7	5.9	24.8	17.5	14.0	27.2	53.7	42.4	21.0	61.6
			200	6.9	12.2	6.1	6.9	63.6	52.3	39.1	63.8	93.1	92.0	73.1	96.7
	0.7	2	50	7.1	19.2	5.9	9.1	10.5	8.7	11.9	7.9	16.6	15.5	16.7	17.5
			100	6.8	13.7	6.5	7.9	16.0	13.6	13.3	17.3	40.5	42.6	38.6	52.9
			200	4.9	11.3	5.0	4.8	62.4	63.0	60.3	70.9	94.1	97.4	93.2	98.5
		5	50	9.0	20.5	7.5	9.2	9.3	6.3	6.8	9.1	16.6	13.6	7.4	18.4
			100	7.5	17.9	8.0	7.7	17.7	14.3	10.2	19.0	51.5	38.1	19.7	57.3
			200	5.6	11.4	4.8	5.4	64.8	56.5	48.1	67.9	93.5	95.3	81.8	98.5
0.5	0.3	2	50	6.2	13.0	5.6	7.9	9.2	7.5	10.2	9.5	16.9	16.8	18.5	17.6
			100	4.7	9.1	4.9	5.8	22.5	21.4	21.4	23.7	52.9	53.6	51.2	53.6
			200	6.4	10.4	6.7	7.7	54.1	51.7	49.9	54.7	91.4	93.0	90.1	93.8
		5	50	7.3	11.5	3.7	8.0	8.4	9.7	11.0	8.6	15.6	16.1	18.5	15.2
			100	5.3	9.9	5.4	5.9	22.9	21.0	18.0	22.3	51.4	46.8	36.6	51.3
			200	5.6	9.4	4.5	6.1	62.2	57.1	58.4	61.8	95.6	93.5	90.0	95.8
	0.7	2	50	5.6	12.8	5.5	8.5	8.8	7.8	9.2	8.1	17.9	16.5	18.6	16.7
			100	5.3	9.7	5.9	6.3	18.8	17.9	17.0	20.7	48.5	46.6	45.2	52.6
			200	5.3	9.2	5.2	6.1	60.1	62.1	60.5	60.9	94.5	96.0	94.8	95.6
		5	50	7.1	12.3	4.4	8.3	8.6	10.1	9.1	9.4	15.5	17.7	13.8	16.3
			100	5.5	10.1	4.7	5.9	19.2	19.1	17.6	21.0	51.0	46.8	41.2	53.7
			200	4.8	9.7	4.8	4.8	63.8	57.8	54.2	67.1	93.7	92.2	88.6	95.8

Notes: The value γ refers to the serial correlation parameter, λ refers to the location of the break, δ refers to size of the break and ϕ refers to the autoregressive parameter. Also, $\widehat{\tau}_\phi$ refers to τ_ϕ based on an estimated break, LM_τ refers to the Lee and Strazicich (2003) minimum test and $\widetilde{\tau}$ refers to the Schmidt and Phillips (1992) no-break test.

Table 2: Size and power for an unknown number of breaks.

				$\phi = 0$		$\phi = -0.1$		$\phi = -0.2$	
K	γ	δ	T	$\widehat{\tau}_\phi$	τ_ϕ	$\widehat{\tau}_\phi$	τ_ϕ	$\widehat{\tau}_\phi$	τ_ϕ
1	0	2	50	5.7	6.4	9.9	11.5	18.6	23.9
			100	6.4	6.6	19.1	21.0	44.5	59.1
			200	4.8	4.1	63.2	70.6	92.7	98.6
		5	50	10.3	10.3	9.4	10.3	16.0	17.8
			100	5.0	4.7	25.4	28.2	58.5	66.5
			200	6.0	5.6	63.3	66.5	93.1	97.9
	0.5	2	50	6.8	8.9	8.2	8.9	13.5	17.5
			100	5.2	6.7	19.2	19.2	46.6	48.6
			200	5.4	5.7	56.3	59.5	93.3	95.8
		5	50	8.3	9.5	6.6	7.0	13.7	14.0
			100	6.4	6.9	20.0	20.2	47.6	49.0
			200	6.2	6.5	59.8	62.3	93.7	95.9
2	0	2	50	9.6	10.0	6.8	7.9	14.1	18.4
			100	6.0	6.7	20.4	21.9	47.4	55.9
			200	5.4	5.6	58.6	65.4	91.6	96.7
		5	50	8.6	8.5	8.5	8.5	15.3	19.2
			100	5.9	5.5	21.9	25.3	51.2	62.5
			200	5.4	5.4	61.3	66.4	89.9	97.2
	0.5	2	50	5.6	9.0	9.0	8.6	18.1	16.8
			100	6.1	7.4	18.0	19.4	43.7	46.0
			200	5.3	6.6	57.0	59.2	92.3	96.0
		5	50	6.1	7.9	8.6	10.2	17.3	18.5
			100	5.6	6.3	22.2	23.2	47.6	51.1
			200	5.3	5.5	53.8	57.0	92.2	96.1

Notes: The value K refers to the number of breaks in the data generating process. See Table 1 for an explanation of the remaining parameters.

Table 3: Breakpoint estimation accuracy.

			$K=1$ known		$K=1$ unknown		$K=2$ unknown	
ϕ	γ	T	$\delta=2$	$\delta=5$	$\delta=2$	$\delta=5$	$\delta=2$	$\delta=5$
0	0	50	28.1	92.2	31.6	87.8	9.9	78.1
		100	28.7	90.7	29.8	86.4	8.9	77.4
		200	27.8	89.9	28.8	86.9	9.6	79.7
	0.5	50	21.3	76.3	22.3	77.4	3.2	56.9
		100	17.6	74.6	19.8	74.5	4.7	58.0
		200	16.0	75.9	18.3	72.5	4.7	57.1
-0.1	0	50	26.5	90.2	23.2	85.7	8.0	74.2
		100	28.1	88.2	24.5	85.7	7.4	76.8
		200	23.7	90.0	26.7	84.9	8.1	73.4
	0.5	50	21.8	79.8	21.5	75.5	5.0	58.4
		100	18.9	77.6	19.9	74.9	4.4	58.1
		200	19.5	77.1	20.1	76.9	5.4	60.8
-0.2	0	50	23.6	86.1	26.7	82.9	6.4	67.5
		100	22.2	85.7	22.6	82.1	7.2	71.1
		200	21.2	87.0	27.7	81.5	6.7	71.7
	0.5	50	22.7	78.6	23.1	78.3	4.3	59.7
		100	18.4	80.3	20.9	77.8	4.0	63.3
		200	19.9	80.1	20.0	77.0	4.8	62.4

Notes: The numbers reported in the table are the percentage of times when the correct breakpoints are selected. See Table 1 for an explanation of the various parameters in the table.

Table 4: Size and power for the panel test.

				$K=1$ known		$K=1$ unknown		$K=2$ unknown	
ϕ	δ	N	T	$\widehat{\tau}_{\phi N}$	$\tau_{\phi N}$	$\widehat{\tau}_{\phi N}$	$\tau_{\phi N}$	$\widehat{\tau}_{\phi N}$	$\tau_{\phi N}$
1	2	5	100	7.3	8.3	8.8	8.8	8.7	10.4
		10	100	11.6	11.2	10.0	10.2	9.1	9.8
		5	200	6.0	6.4	7.0	7.3	7.7	8.2
		10	200	6.6	6.8	7.3	7.5	5.8	5.9
	5	5	100	8.1	8.2	8.3	8.3	9.2	9.4
		10	100	10.5	10.6	10.4	9.7	10.7	11.3
		5	200	7.1	7.1	7.1	7.1	6.6	7.0
		10	200	7.1	6.9	9.7	9.3	8.6	8.8
0.95	2	5	100	27.2	28.2	20.8	21.7	21.4	26.5
		10	100	39.2	42.4	34.1	39.8	36.8	42.1
		5	200	80.0	83.8	81.8	84.6	77.4	81.5
		10	200	97.9	98.7	97.0	98.7	97.2	99.1
	5	5	100	27.4	28.2	26.5	28.5	20.4	22.0
		10	100	39.0	42.2	41.6	42.7	35.7	38.9
		5	200	81.7	82.3	80.5	82.8	80.8	83.5
		10	200	98.6	98.8	96.1	96.7	96.8	97.0
0.9	2	5	100	72.4	78.5	68.6	73.9	63.0	76.2
		10	100	92.5	96.9	93.2	97.2	91.6	96.8
		5	200	100.0	100.0	100.0	100.0	99.7	99.9
		10	200	100.0	100.0	100.0	100.0	100.0	100.0
	5	5	100	77.5	80.3	76.9	79.8	68.2	76.1
		10	100	96.1	97.7	97.5	98.4	93.1	96.0
		5	200	100.0	100.0	100.0	100.0	100.0	100.0
		10	200	100.0	100.0	100.0	100.0	100.0	100.0

Notes: The $\widehat{\tau}_{\phi N}$ test refers to $\tau_{\phi N}$ based on estimated breaks. See Table 1 for an explanation of the various parameters in the table.

Table 5: Country-specific unit root tests.

	Unit root test				Estimated breaks	
Country	$\widehat{\tau}_{\phi}$	p-val	$\widetilde{\tau}$	p-val	No.	Location
Australia	−2.019	0.424	−2.019	0.424	0	–
Austria	−2.441	0.204	−2.441	0.204	0	–
Belgium	−1.745	0.610	−1.745	0.610	0	–
Canada	−1.485	0.788	−1.485	0.788	0	–
Denmark	−1.415	0.833	−1.415	0.833	0	–
Finland	−2.117	0.363	−2.117	0.363	0	–
France	−1.568	0.740	−1.442	0.815	1	1978:Q4
Germany	−2.438	0.206	−2.438	0.206	0	–
Greece	−0.931	0.993	−1.016	0.984	1	1995:Q1
Iceland	−1.241	0.920	−1.151	0.954	1	1995:Q1
Ireland	−1.393	0.845	−1.393	0.845	0	–
Italy	−1.978	0.452	−1.825	0.555	1	1995:Q1
Japan	−1.694	0.646	−1.694	0.646	0	–
Luxembourg	−1.842	0.542	−1.842	0.542	0	–
Netherlands	−2.172	0.331	−2.172	0.331	0	–
Norway	−1.653	0.680	−2.064	0.393	1	1995:Q1
Portugal	−1.571	0.737	−1.688	0.651	1	1995:Q1
Spain	−1.826	0.554	−1.826	0.554	0	–
Sweden	−1.616	0.706	−2.119	0.362	1	1995:Q1
Switzerland	−1.739	0.614	−1.739	0.614	0	–
United Kingdom	−1.929	0.486	−1.929	0.486	0	–

Notes: The $\widehat{\tau}_{\phi}$ test refers to τ_{ϕ} based on estimated breaks, while $\widetilde{\tau}$ refers to the Schmidt and Phillips (1992) no-break test. The p-values are based on simulating the asymptotic distribution of the test. This is done using $10,000$ random walks of length $1,000$.

Table 6: Panel unit root tests.

Panel	N	$\widehat{\tau}_{\phi N}$	p-val[a]	p-val[b]	$\widetilde{\tau}_{N}$	p-val[a]	p-val[b]
Choi (2001)	6	0.758	0.743	0.776	0.918	0.782	0.821
European monetary union	12	0.817	0.674	0.793	0.855	0.684	0.804
Europen community	15	1.005	0.679	0.843	0.811	0.646	0.791
OECD	21	1.742	0.784	0.959	1.455	0.744	0.927

Notes: The $\widehat{\tau}_{\phi N}$ test refers to $\tau_{\phi N}$ based on estimated breaks, while $\widetilde{\tau}_{N}$ refers to the panel version of the Schmidt and Phillips (1992) no-break $\widetilde{\tau}$ test.
[a]The p-values are based on the bootstrapped distribution using $1,000$ replications.
[b]The p-values are based on the asymptotic normal distribution.

References

Ahn, S. K. (1993). Some Tests for Unit Roots in Autoregressive-Integrated-Moving Average Models with Deterministic Trends. *Biometrica*, Vol. 80, pp. 855-868.

Amsler, C., and J. Lee (1995). An LM Test for a Unit Root in the Presence of a Structural Break. *Econometric Theory*, Vol. 11, pp. 359-368.

Chang, Y. (2004). Bootstrap Unit Root Tests in Panels with Cross-Sectional Dependency. *Journal of Econometrics*, Vol. 120, pp. 263-293.

Choi, I. (2001). Unit Root Tests for Panel Data, *Journal of International Money and Finance*, Vol. 20, pp. 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*, Vol. 23, pp. 1159-1186.

Craner, M., and L. Kilian (2001). Size Distortions of Tests of the Null Hypothesis of Stationarity: Evidence and Implications for the PPP Debate. *Journal of International Money and Finance*, Vol. 20, pp. 639-657.

Gadea, M-D., A. Montañés and M. Reyes (2004). The European Union Currencies and the US Dollar: From Post-Bretton-Woods to the Euro. *Journal of International Money and Finance*, Vol. 23, pp. 1109-1136.

Im, K. S., J. Lee, and M. Tieslau (2005). Panel LM Unit Root Tests with Level Shifts. *Oxford Bulletin of Economics and Statistics*, Vol. 67, pp. 393-419.

Kapetanios, G. (2005). Unit-Root Testing Against the Alternative Hypothesis of up to m Structural Breaks. *Journal of Time Series Analysis*, Vol. 26, pp. 123-133.

Lee, J., and M. Strazicich (2003). Minimum LM Unit Root Test with Two Structural Breaks. *Review of Economics and Statistics*, Vol. 85, pp. 1082-1089.

Papell, D. H., and H. Theodoridis (1998). Increasing Evidence of Purchasing Power Parity Over the Current Float. *Journal of International Economics*, Vol. 17, pp. 41-50.

Perron, P. (1989). The Great Crash, the Oil Price Shock and the Unit Root Hypothesis. *Econometrica*, Vol. 57, pp. 1361-1401.

Perron, P., and G. Rodríguez (2003). Searching for Additive Outliers in Nonstationary Time Series. *Journal of Time Series Analysis*, Vol. 24, pp. 193-220.

Sabatè, M., M-D. Gadea and J. M. Serrano (2003). PPP and Structural Breaks. The Peseta−Sterling Rate, 50 Years of a Floating Regime. *Journal of International Economics*, Vol. 22, pp. 613-627.

Vogelsang, T. J. (1999). Two Simple Procedures for Testing for a Unit Root when there are Additive Outliers. *Journal of Time Series Analysis*, Vol. 20, pp. 237-252.

Zivot, E., and D. W. K. Andrews (1992). Further Evidence on the Great Crash, the Oil-Price Shock and the Unit-Root Hypothesis. *Journal of Business and Economics Statistics*, Vol. 10, pp. 251-270.