Financial Econometrics Series

SWP 2011/07

Size and Power Properties of Structural Break Unit Root Tests

P.K. Narayan and S. Popp

The working papers are a series of manuscripts in their draft form. Please do not quote without obtaining the author’s consent as these works are in their draft form. The views expressed in this paper are those of the author and not necessarily endorsed by the School or IBISWorld Pty Ltd.
Size and Power Properties of Structural Break Unit Root Tests

Paresh Kumar Narayan (Deakin University)
and Stephan Popp (University of Duisburg-Essen)

August 5, 2011

Abstract

In this paper, we compare the small sample size and power properties of a newly developed endogenous structural break unit root test of Narayan and Popp (NP, 2010) with existing two break unit root tests, namely the Lumsdaine and Papell (LP, 1997) and the Lee and Strazicich (LS, 2003) tests. In contrast to the widely used LP and LS tests, the NP test chooses the break date by maximising the significance of the break dummy coefficient. Using Monte Carlo simulations, we show that the NP test has better size and high power, and identifies the structural breaks accurately. Power and size comparisons of the NP test with the LP and LS tests reveal that the NP test is significantly superior.
1 Introduction

In applied time series, testing for the integrational properties of the data series is a pre-requisite for modelling relationships. The importance of unit root testing has also close links with economic theory. In the applied economics literature, there are several tests of the unit root hypothesis relating to purchasing power parity (see, inter alia, Grilli and Kaminsky, 1991; Narayan, 2005, 2006), business cycles (see, inter alia, Cheung and Chinn, 1996; Cogley, 1990; Kormendi and Meguire, 1990), and the intertemporal model of the current account (see, inter alia, Lau and Baharumshah, 2005; Wu, 2000), among others. It follows that the motivation for the advent of new unit root tests with better statistical properties has strong roots in economic theory.

Perron’s (1989) exogenous unit root test, which allows for one structural break, has been the most widely used. Zivot and Andrews (ZA, 1992) modified and extended the Perron test to a case of an endogeneous structural break. Subsequently, because of its endogeneous treatment of the structural break, the ZA test became popular. The ZA test was further extended to allow for two endogenous structural breaks by Lumsdaine and Papell (LP, 1997) and by Lee and Strazicich (LS, 2003a). It is observed that in applied research, where a sufficiently long time series data is used, the LP and the LS tests have become more popular. The most recent contribution to this
literature is Narayan and Popp (NP, 2010), who differ from LP and LS on the one hand in using a Dickey-Fuller-type test regression and on the other hand in their treatment of break date selection. They select the break dates by maximising the significance of the break dummy coefficient. NP show that their test is invariant to the break magnitude and detects the break date with higher precision. A detailed overview of the various tests are provided in the next section.

The aim of this paper is to undertake a comparative analysis of the finite sample size and power properties of the two widely used structural break unit root tests (the LS and LP tests) with the newly developed test of NP. This comparison will be useful for applied researchers in identifying which test to use for their work. At this point, it should be noted that the bulk of the applied work that considers the unit root hypothesis relates to sample sizes of less than 100; hence, our simulation exercise for comparison is based on a sample size of 100 observations. An analysis of the size and power properties of the NP test in infinite samples is available in NP (2010).

We organise the balance of the paper as follows. In section 2, we give an overview of the existing unit root tests that account for two structural breaks. In section 3, we compare the size and power properties of the most commonly used tests in the applied economics literature. In the final section, we provide some concluding remarks.
2 An overview of existing unit root tests with two breaks

There are two unit root tests, namely the LP (1997) test and the LS (2003) test that are widely used in the applied economics literature. These tests are all flexible enough to allow for at most two structural breaks. The aim of this section is to provide a brief overview of the key features of these two tests. We also briefly introduce the NP (2010) test here as well.

The LP test is a generalisation of the one endogenous structural break unit root test of Banerjee et al. (1992) and ZA (1992). It allows for two endogenous structural breaks in the level and trend (Model 2, M2). The test model M2 is of the form, while model M1 (breaks in level of trending data) is one without the DT terms:

\[
y_t = \alpha + \beta t + \theta_1 DU_{1,t} + \gamma_1 DT_{1,t} + \theta_2 DU_{2,t} + \\
+ \gamma_2 DT_{2,t} + \rho y_{t-1} + \sum_{j=1}^{k} \beta_j \Delta y_{t-j} + e_t \quad (1)
\]

with \( DU_{i,t} = 1(t > T_{B,i}) \) and \( DT_{i,t} = 1(t > T_{B,i})(t - T_{B,i}), i = 1, 2, 1(.) \) denoting the indicator function, and \( T_{B,i} \) is the timing of the break date. The unit root hypothesis is tested using the t-value of \( \rho \), denoted \( t_\rho \). The break date is selected using the break selection \( \arg \min_{T_B} t_\rho(T_B) \) criteria. A
key limitation of the LP type test is noted by LS (2003:1082), who state "One important issue common to the ZA and LP (and other similar) endogenous break tests is that they assume no break(s) under the unit root null and derive their critical values accordingly. Thus, the alternative hypothesis would be 'structural breaks are present', which includes the possibility of a unit root with break(s). Thus, rejection of the null does not necessarily imply rejection of a unit root per se, but would imply rejection of a unit root without breaks.”

The main property of the LP test is that it is not invariant to break size under the null, leading to spurious rejections of the null if a break is present (as is the case for the one break case of ZA). Nunes, Newbold, and Kuan (1997) and Lee and Strazicich (2001) provide evidence that assuming no break under the null in endogenous tests causes the test statistic to diverge and lead to significant rejections of the unit root null when the data generating process (DGP) is a unit root with break(s).

The LS test, on the other hand, is a generalisation of the Schmidt and Phillips (1992) and Lee and Strazicich (2004) tests and allows for breaks under the null and alternative hypotheses for trending data. Their test allows for two endogenous breaks in the level and trend, and begins with the following DGP:

\[ y_t = \Phi'X_t + e_t, \quad e_t = \beta e_{t-1} + \varepsilon_t \]  

(2)
where $X_t$ is a vector of exogenous variables and $\varepsilon_t \sim iid N(0, \sigma^2)$. Model M1, which includes two breaks in the level, has $X_t = [1, t, DU_{1,t}, DU_{2,t}]'$. Model 2, which allows two breaks in the level and slope, is represented by $X_t = [1, t, DU_{1,t}, DU_{2,t}, DT_{1,t}, DT_{2,t}]'$.

The two-break LM unit root test statistic is estimated as follows:

$$\Delta y_t = \Phi' \Delta X_t + \phi \tilde{S}_{t-1} + \mu_t,$$  \hspace{1cm} (3)

where, for model M2, for example, $\tilde{S}_t = y_t - \tilde{\alpha}_u - \tilde{\beta}t - \tilde{\theta}_1 DU_{1,t} - \tilde{\theta}_2 DU_{2,t} - \tilde{\gamma}_1 DT_{1,t} - \tilde{\gamma}_2 DT_{2,t}$, $t = 2, \ldots, T$; $\tilde{\beta}$, $\tilde{\theta}_i$ and $\tilde{\gamma}_i$ are coefficients in the regression of $\Delta y_t$ on a constant, $D(T_B)_i,t$ and $DU_{i,t}$, $i = 1, 2$; $\tilde{\alpha}_u$ is the restricted Maximum Likelihood estimate of $\alpha_u$; see LS (2003:1083). For M1 the dummy variables $DT_{i,t}$ are excluded. The two-break LM unit root test statistic $t^*_\phi,LM$ is used to test the null hypothesis $\rho = 1$ or equivalently $\phi = \rho - 1 = 0$ in Equation (3).

The break date selection is based on the $\arg\min_{T_B} t^*_{\phi,LM}(T_B)$ method by using a grid search. The main property of their test is that it is asymptotically invariant to break size and timing of the break under $H_0$ for M1 (i.e. no spurious rejections), but not exactly invariant to the break location for M2. However, in no case does the LM test diverge or exhibit any systematic pattern of overrejections in the presence of breaks under the null, as noted...
The test proposed by Narayan and Popp (2010) generalises the one break unit root test by Popp (2008a). The IO-type test allows for structural breaks under the null and the alternative hypothesis and is applicable for all model specifications, i.e. M1 and M2. The model is formulated as follows:

\[ y_t = \alpha + \beta t + \delta_1 D(T_B)_{1,t} + \delta_2 D(T_B)_{2,t} + \theta_1 DU_{1,t-1} + \theta_2 DU_{2,t-1} + \gamma_1 DT_{1,t-1} + \gamma_2 DT_{2,t-1} + \rho y_{t-1} + \sum_{j=1}^{\infty} \beta_j \Delta y_{t-j} + e_t. \] (4)

The model M1 can be derived by setting: \( \gamma_1 = \gamma_2 = 0 \). The \( t \)-statistic of \( \hat{\rho} \), \( t_{\hat{\rho}} \), serves as the test statistic to examine the null hypothesis that \( \rho = 1 \). The break dates can be identified using a grid search by using the following selection method: \( \arg \max_{T_B} F_{\theta_1, \theta_2}(T_B) \). To minimise the computational burden, NP adopt a sequential approach. The method described in NP selects the break date by maximizing the significance of the break dummy coefficients: \( \arg \max_{T_B} |t_{\hat{\delta}_i}(T_B,i)| \). The test has the following properties: (1) it is invariant to the break magnitude even in finite samples, avoiding the problem of spurious rejections when a break is present under \( H_0 \); and (2) the break date is detected very accurately.
3 Finite sample size, power and break date estimation accuracy

The finite sample size and power is assessed by simulation analysis. We assume the following model:

\[
y_t = d_t + u_t, \quad u_t = \rho u_{t-1} + \epsilon_t, \tag{5}
\]

\[
d_t = \alpha + \beta t + \theta_1 DU_{1,t} + \gamma_1 DT_{1,t} + \theta_2 DU_{2,t} + \gamma_2 DT_{2,t}, \tag{6}
\]

where \( \epsilon_t \sim iid N(0,1) \). The dummy variables are defined as earlier. The parameters \( \theta_i \) and \( \gamma_i, i = 1,2 \), denote the level and slope break size, respectively. The results are based on 5000 replications of a sample of \( T = 100 \) each. We generated 150 observations and discarded the first 50 observations in order to avoid any effects of the initial conditions. The break dates are assumed to be \( T_B = (T_{B,1} = 40, T_{B,2} = 60) \). In our preliminary analysis, we also considered several other break date combinations and found similar results. In order to conserve space, here we only report results for this one case. The rest of the results are available from the authors upon request.

We further assume for M1 that there exists two periods, and for M2 three periods, between the first and second break date: \( |T_{B,2} - T_{B,1}| \geq c \), where \( c = 2 \) for M1 and \( c = 3 \) for M2. The trimming factor is \( \tau = 0.2 \). We con-
duct the simulations using GAUSS 8.0. The empirical size and power are computed for $\rho = 1$ and $\rho = 0.9$, respectively. Furthermore, we vary the magnitude of level and slope break to assess the ability of the test to detect the true break date and to verify the break size invariance property in finite samples. We assume the break magnitudes to be identical for both breaks, i.e. $\theta = \theta_1 = \theta_2$ and $\gamma = \gamma_1 = \gamma_2$. Although not considered in this paper for purposes of tractability, a more general autocorrelation structure can also be assumed.

We analyse the test properties of LP and LS tests using $\text{arg min } t_{\rho}$ and $\text{arg min } t_{\hat{\phi},LM}$, respectively. Moreover, we consider the test by NP (2010). For test decision, the critical values derived under the assumption of no break, that is, a break size of zero is used.

We also judge the test properties of LS and NP using the critical values of the respective tests when the break dates are exogenously given. These tests are denoted LSCVexo and NPCVexo, respectively. We use the set of critical values for known break dates because if with increasing break size the probability of detecting the true break date goes to 1, $\lim_{\text{break size} \to \infty} P(\hat{T}_B = T'_B) = 1$ and additionally the test is invariant to the break size, it implies that the critical values (and distribution) for the endogenous test assuming

---

1 Thanks to Junsoo Lee for providing the GAUSS codes of the LP and LS tests.

2 A situation in which one always identifies the break date correctly, i.e. $P(\hat{T}_B = T'_B) = 1$, is like knowing the break date.
no break (denoted CVendo) is equivalent to that for the exogenous test (denoted CVexo). When the difference between CVendo and CVexo is large and the break date estimation accuracy increases with the break size, this leads to tests with unstable size. An overview of the various tests and their denotations are provided in Table 1.

3.1 Size effects

For the M1 test, the empirical size for the NP test is slightly undersized but relatively close to the nominal 5 per cent level regardless of the size of the break. The LS test is substantially undersized for large breaks (see Figure 1). The reason for this is that the test distribution converges with the break size to that of the exogenous break test. By comparison, the ADF test is relatively more undersized while the LP test is substantially oversized.

When we use the critical values of the exogenous break test, the size converges to the 5 per cent nominal level for NPCVexo. In the case of LSCVexo, however, the size of the small breaks is between 12-28 per cent. With increas-

\footnote{The Perron test is a Dickey-Fuller type test. As stated by Perron (1988), the exogenous break test is invariant to the break date if we account for the break. In case of an unknown break date, the test is invariant if it is possible to identify the break date accurately. So, for an endogenous break unit root test, it is important to meet these preconditions. The advantage of the NP test is that it identifies the break date very accurately, even in the case of very small breaks. The reason for this is the slightly different approach for estimating the break date. One indicator of good properties of of the NP test is that the critical values of the endogenous and exogenous tests are very similar in finite samples, which is not the case with the other tests considered in this paper.}
ing break size, one can observe our earlier claim that there is convergence of the test distribution of the endogenous break test to that of the exogenous break test. This is shown by the convergence of the empirical size to their nominal value of 5 per cent. Moreover, the ability of the NP test to identify both breaks simultaneously is high with large breaks compared with the LP and LS tests (see Figure 2). In fact, the LS test performs even poorly than the LP test.

To assess the size properties for M2, we vary the level and slope break size. To avoid three-dimensional figures, we generate two figures, one with fixed level break size and variable slope break size, and vice versa (see Figures 3 and 4). For the M2 test, the empirical size of the NP and NPCVexo test is around the nominal 5 per cent level for both level breaks (Figure 3) and slope breaks (Figure 4). By comparison, the LS and LSCVexo tests with two endogenous structural breaks diverge from the 5 per cent nominal level both with increasing level and slope break magnitude. Especially for the realistic case of small slope breaks, the LSCVexo test has size of around 20-30 per cent.

The probability of detecting small to large sized breaks is close to 100 per cent for the NP test, suggesting that the model detects breaks accurately for the case of both level break and slope break (see Figures 5 and 6). This performance of the NP test is superior compared with the LP and LS two
break tests.

We report the size properties of the LP test for model M1 in Figure 1 and for model M2 in Figures 3 and 4. We find that the empirical size diverges substantially from the 5 per cent nominal level for medium to large sized breaks and the test has weak ability to detect the correct level breaks simultaneously, as depicted in Figures 2 and 5. For M2, the LP test performs relatively poorly. The empirical size for medium to large sized breaks are close to 100 per cent (see Figures 3 and 4). Furthermore, the test is unable to detect the correct dates of slope breaks (see Figure 6).

3.2 Empirical power

The empirical power of the NP and LS tests for model M1 are reported in Figure 7. We also report the power properties of the conventional ADF test for the sake of comparison. As is well known, the ADF test has low power against the unit root hypothesis in the presence of breaks; we find this to be the case as well. However, we do not report the power properties of the LP test because it is significantly oversized, as we found earlier. For a test with such high size distortions, it is meaningless to estimate power properties.

We find that the power of the NP test is stable, reflecting its stable size, and it is invariant to the size of the break, while the power of the LS test
decreases with the break magnitude, and is lower than the NP test (M1). However, when we use the CVs of the known break dates, the power of LS is higher than NP, though the power declines with increasing break size. In terms of break date estimation accuracy, as reported in Figure 8, we find that the power of the LS test to detect the break is close to 100 per cent for medium to large breaks.

In Figures 9 and 10 we report the empirical power of the M2 test for the level break and slope break, respectively. The power of LS and LSCVexo increases with the magnitude of the breaks which is due to their diverging size. The power is found to be invariant to the break size for the NP test. The NP test is also able to identify the correct breaks simultaneously: 100 per cent accuracy for both medium to large sized breaks (see Figures 11 and 12). By comparison, the power results for the LS test based on \( \hat{T}_B = \arg\min_{\hat{\phi}_{LM}} \) suggest that it is slightly higher than the NP test. However, the NP test is still invariant to the break size and the ability of the LS test to detect the correct break dates simultaneously is close to zero for medium to large sized breaks.
3.3 Summary

Upon comparing the size and power properties of the NP test with the existing tests for unit roots with two structural breaks, we observe the following. For M1, the LP test reveals that the 5 percent rejection frequency approaches 100 per cent with increasing break size. But, the test identifies the break dates more accurately with increasing break size. For M2, the 5 percent rejection frequency converges to 100 percent more rapidly. Furthermore, the test is weak in detecting the level break dates for small and medium sized breaks and unable to detect slope break dates.

In the case of M1, for the LS test, the 5 percent empirical size is very stable for small break, but gets very conservative for medium and large breaks. The test hardly detects the break dates with increasing break size. Because the frequency of detecting the true break dates increases very slowly, the mentioned convergence from the endogenous test distribution to the exogenous test distribution takes place slowly, giving the impression of an invariant test with stable size. But for very large breaks the test is conservative.

The performance of the LS test for M2 is relatively weak. The test has difficulties in identifying the break dates. Moreover, the test using the CV for no break (CVendo) and for given break date (CVexo) is largely distorted.

Finally, the NP test consists of the following features. First, for model
M1, the performance of detecting the break dates is significantly better than the LS test. Using the CVendo, the test gets a little bit conservative. Using the CVexo leads to small oversizing for small breaks and the empirical size converges to the nominal size for medium and large breaks. So, NP’s M1 model has empirical size always close to the nominal one as opposed to the LS test.

The M2 test of NP has the best performance in terms of detecting the break dates. Although the test is slightly conservative for CVendo and slightly oversized for CVexo in the case of small break sizes, it performs better than the existing tests.

The main advantage of the NP test is that it unites many favourable properties in terms of stable size in the presence of level and slope breaks, and accurate estimation of the break dates under both the null and alternative hypotheses. So, the applied economist has with the NP test a test procedure at hand that gives relatively good results for all model specifications.

4 Concluding remarks

In this paper, we compared the finite sample size and power properties of the newly developed two structural break unit root test of Narayan and Popp (NP, 2010) with two existing and widely used unit root tests of Lumsdaine
and Papell (LP, 1997) and Lee and Strazicich (LS, 2003). The key feature of the NP test that distinguishes it from the LP and LS tests is that it uses a Dickey-Fuller-type test approach and it chooses the break date by maximising the significance of the break date coefficient.

Using Monte Carlo simulations, we show that the NP test has good size and stable power, and identifies the structural breaks accurately in finite samples. Power and size comparisons of the NP test with the most widely used two structural break test of LP and LS reveals that the NP test is significantly superior.
References


<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADF</td>
<td>ADF test</td>
</tr>
<tr>
<td>LP</td>
<td>procedure according to Lumsdaine and Papell (1997), arg min $t_\hat{\rho}$</td>
</tr>
<tr>
<td>LS</td>
<td>procedure according to Lee and Strazicich (2003), arg min $t_{\hat{\rho},LM}$</td>
</tr>
<tr>
<td>LSexo</td>
<td>procedure according to Lee and Strazicich (2003), known break date</td>
</tr>
<tr>
<td>LSCVexo</td>
<td>LS using the critical values from LSexo</td>
</tr>
<tr>
<td>NP</td>
<td>procedure according to Narayan and Popp (2009), sequential</td>
</tr>
<tr>
<td>NPexo</td>
<td>procedure according to Narayan and Popp (2009), known break date</td>
</tr>
<tr>
<td>NPCVexo</td>
<td>NP using the critical values from NPexo</td>
</tr>
</tbody>
</table>

Figure 1: 5% rejection frequency under $H_0$, $T = 100$, $T_B = (T_{B,1} = 40, T_{B,2} = 60)$, M1
Figure 2: Break date estimation accuracy under $H_0$, $T = 100$, $T_B = (T_{B,1} = 40, T_{B,2} = 60)$, M1
Figure 3: 5% rejection frequency under $H_0$, $T = 100$, $T_B = (T_{B,1} = 40, T_{B,2} = 60)$, fixed slope break size $\gamma = 2$, varying level break size $\theta$, M2
Figure 4: 5% rejection frequency under $H_0$, $T = 100$, $T_B = (T_{B,1} = 40, T_{B,2} = 60)$, fixed level break size $\theta = 2$, varying slope break size $\gamma$, M2
Figure 5: Break date estimation accuracy under $H_0$, $T = 100$, $T_B = (T_{B,1} = 40, T_{B,2} = 60)$, fixed slope break size $\gamma = 2$, varying level break size $\theta$, M2
Figure 6: Break date estimation accuracy under $H_0$, $T = 100$, $T_B = (T_{B,1} = 40, T_{B,2} = 60)$, fixed level break size $\theta = 2$, varying slope break size $\gamma$, M2
Figure 7: 5% rejection frequency under $H_1 (\rho = 0.9), T = 100, T_B = (T_{B,1} = 40, T_{B,2} = 60)$, M1
Figure 8: Break date estimation accuracy under $H_1 (\rho = 0.9)$, $T = 100$, $T_B = (T_{B,1} = 40, T_{B,2} = 60)$, M1
Figure 9: 5% rejection frequency under $H_1 (\rho = 0.9)$, $T = 100$, $T_B = (T_{B,1} = 40, T_{B,2} = 60)$, fixed slope break size $\gamma = 2$, varying level break size $\theta$, M2
Figure 10: 5% rejection frequency under $H_1 (\rho = 0.9), T = 100, T_B = (T_{B,1} = 40, T_{B,2} = 60)$, fixed level break size $\theta = 2$, varying slope break size $\gamma$, M2
Figure 11: Break date estimation accuracy under $H_1$ ($\rho = 0.9$), $T = 100$, $T_B = (T_{B,1} = 40, T_{B,2} = 60)$, fixed slope break size $\gamma = 2$, varying level break size $\theta$, M2
Figure 12: Break date estimation accuracy under $H_1$ ($\rho = 0.9$), $T = 100$, $T_B = (T_{B,1} = 40, T_{B,2} = 60)$, fixed level break size $\theta = 2$, varying slope break size $\gamma$, M2