Bounds, breaks and unit root tests[*]

Josep Lluís Carrion-i-Silvestre[†]   María Dolores Gadea[‡]

University of Barcelona   University of Zaragoza

May 8, 2015

Abstract

The paper addresses the unit root testing when the range of the time series is limited and considering the presence of multiple structural breaks. The structural breaks can affect the level and/or the boundaries of the time series. The paper proposes five unit root test statistics, whose limiting distribution is shown to depend on the number and position of the structural breaks. The performance of the statistics is investigated by means of Monte Carlo simulations.

Keywords: multiple structural breaks, regulated Brownian motion, limited time series, unit root

JEL codes: C12, C22

1 Introduction

The application of standard unit root and stationarity tests assumes that time series are not limited or bounded – i.e., its range of variation is not constrained. However, there are some situations where time series are bounded below, above, or both, so their range of variation is restricted, and defines limited stochastic processes. Composition variables such as percentage of expenditure on different concepts, unemployment rates, participation in labor market ratios, income inequality measures, capacity utilization ratios and number of hours worked are some examples of time series that are limited either by construction or by definition. There are also variables, such as nominal interest rates, that cannot take negative values, so they are bounded below. Finally, variables such as the exchange rates of the currencies within the European Monetary System (EMS) have been controlled and move inside bands of fluctuation.

Cavaliere (2005) and Cavaliere and Xu (2014) show that the limit distributions of the Dickey-Fuller (DF) and the modified unit root tests in Ng and Perron (2001) – henceforth, M tests – shift to the left when a bounded time series takes values close to the boundaries, which causes severe overrejection distortions of the null hypothesis of unit root for the standard unit root tests. Intuitively, bounded I(1) processes – hereafter, BI(1) – might be characterized as stationary I(0) processes because of the bounds, since the bounds prevent the processes from drifting away from the expected value. Consequently, standard unit root and stationarity tests might lead to conclude that a given process is mean-reverting, although mean reversion is due to the bounds, not to the I(0)-ness of the process.

The availability of long time series is a desirable feature in order to ensure the application of unit root tests with good properties. However, the longer the time period, the higher the probability of having structural breaks affecting the behavior of the time series. Perron (1989, 1990) drew attention to the bias shown by the DF statistic towards the non-rejection of the null hypothesis of unit root when structural breaks are present but not accounted for. Practitioners working with limited time series might be facing two misspecification errors that bias the unit root test statistics in opposite directions.

The aim of this paper is to propose unit root tests for limited time series that take into

account the presence of multiple structural breaks. We show that the limiting distribution of the test statistics is affected by the number and position of the structural breaks. There are two different parts of the model where the structural breaks can play a role. First, and following Perron (1989, 1990), the structural breaks can affect the deterministic component of the model. Second, they can also change the boundaries that restrict the variation of the time series. These two sources of parameter instability can be combined, giving rise to different situations that are important from an empirical point of view. This paper covers all these cases and shows that the limiting distribution of the unit root tests depends not only on the number and position of the structural breaks, but also on whether the structural breaks affect the deterministic component and/or the boundaries of the time series.

The paper is structured as follows. Section 2 introduces the model. Section 3 presents the unit root tests statistics for broken bounded time series and derives their limiting distribution. Section 4 deals with the unknown breaks case. The finite sample performance of the unit root statistics is investigated in Section 5. Section 6 illustrates the application of the proposal. Finally, Section 7 concludes. All the proofs are contained in the appendix.

2   The model

Let us define the following data generating process (DGP):

$$x_t = d_t + y_t \tag{1}$$

$$d_t = \mu + \sum_{i=1}^{m} \theta_i DU_{i,t} \tag{2}$$

$$y_t = \alpha y_{t-1} + u_t, \tag{3}$$

$t = 1, \ldots, T$, where $x_t \in \left[\underline{b_t}, \overline{b_t}\right]$ almost surely for all t, $DU_{i,t} = 1$ for $t > T_i^0$ and 0 otherwise, for $i = 1, \ldots, m$ structural breaks – the superscript 0 in T_i^0 is used to denote the true break date – and initialized at $y_0 = O_p(1)$. Further, define $\lambda_i^0 = T_i^0/T$ as the break fraction parameters, $i = 1, \ldots, m$, which are collected in the vector $\lambda^0 = \left(\lambda_1^0, \ldots, \lambda_m^0\right)'$. The presence of bounds requires that Δx_t lies within the interval $\left[\underline{b_t} - x_{t-1}, \overline{b_t} - x_{t-1}\right]$, where $\left[\underline{b_t}, \overline{b_t}\right]$ denote

the boundaries that affect the time series. These boundaries are assumed to be defined by deterministic terms. Cavaliere (2005) focuses on the analysis of non-trending variables defining $\left[\underline{b}_t, \overline{b}_t\right] = \left[\underline{b}, \overline{b}\right]$ $\forall t$. In this paper, we restrict the analysis to non-trending variables and the changing boundaries can be specified as $\underline{b}_t = \underline{b} + \sum_{i=1}^{m} \underline{\gamma}_i \, 1\left(T_{i-1}^0 < t \leq T_i^0\right)$ and $\overline{b}_t = \overline{b} + \sum_{i=1}^{m} \overline{\gamma}_i \, 1\left(T_{i-1}^0 < t \leq T_i^0\right)$, where $1\left(\cdot\right)$ is the indicator function. The disturbance term u_t is:

$$u_t = \varepsilon_t + \underline{\xi}_t - \overline{\xi}_t, \tag{4}$$

where ε_t is a weakly dependent zero-mean process with infinite support and $\underline{\xi}_t$ and $\overline{\xi}_t$ are non-negative processes (regulators) such that $\underline{\xi}_t > 0$ if and only if $y_{t-1} + \varepsilon_t < \underline{b}_t - d_t$ and $\overline{\xi}_t > 0$ if and only if $y_{t-1} + \varepsilon_t > \overline{b}_t - d_t$. The stochastic processes involved in (4) satisfy:

Assumption 1: $\varepsilon_t = C\left(L\right) v_t$, where $C\left(L\right) = \sum_{i=0}^{\infty} C_i L^i$ with $\sum_{i=0}^{\infty} i^s \left|C_i\right| < \infty$ for some $s \geq 1$, and v_t is a martingale difference sequence adapted to the filtration $F_t = \sigma - field\left\{v_{t-i}; i \geq 0\right\}$. The long-run variance (LRV) of ε_t is given by (a) $\sigma^2 = \lim_{T \to \infty} E[T^{-1}(\sum_{t=1}^{T} \varepsilon_t)^2] = \sigma_v^2 C\left(1\right)^2$, (b) $\sigma_v^2 = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E\left(v_t^2\right) < \infty \, \forall t$, and (c) $E\left|v_t^r\right| < \infty$ for some $r > 4$.

Assumption 2: $\left\{\underline{\xi}_t\right\}_{t=1}^{T}$ and $\left\{\overline{\xi}_t\right\}_{t=1}^{T}$ satisfy restrictions to ensure that $\max_{t=1,\ldots,T} \left|\underline{\xi}_t\right| = o_p\left(T^{1/2}\right)$ and $\max_{t=1,\ldots,T} \left|\overline{\xi}_t\right| = o_p\left(T^{1/2}\right)$.

Assumption 3: $\left(\underline{b}_t - d_t\right) = \underline{c}_i \sigma T^{1/2}$ and $\left(\overline{b}_t - d_t\right) = \overline{c}_i \sigma T^{1/2}$ for $T_{i-1}^0 < t \leq T_i^0$, $i = 1, 2, \ldots, m+1$, with $T_0^0 = 0$, $T_{m+1}^0 = T$, and $\underline{c}_i \leq 0 \leq \overline{c}_i$, $\underline{c}_i \neq \overline{c}_i$.

Following Cavaliere (2005), we define general time-dependent lower and upper bounds as:

$$\left[\frac{\left(\underline{b}_t - d_t\right)}{\sigma T^{1/2}}, \frac{\left(\overline{b}_t - d_t\right)}{\sigma T^{1/2}}\right] = \left[\underline{c}_i, \overline{c}_i\right] + o\left(1\right) \text{ for } T_{i-1}^0 < t \leq T_i^0,$$

$i = 1, 2, \ldots, m+1$, with $T_0^0 = 0$ and $T_{m+1}^0 = T$. Note that the time dependence of the bounds $\left(\left[\underline{c}_i, \overline{c}_i\right]\right)$ is introduced through either the boundaries $\left(\left[\underline{b}_t, \overline{b}_t\right]\right)$ and/or the deterministic component $\left(d_t\right)$ of the time series. As a matter of interpretation, it is necessary to stress the different role that plays the boundaries (b's) and the bounds (c's) in this setup. First, in most of cases the boundaries are known by the researcher because they are either implicit

4

by the own definition of the variable – for instance, the unemployment rate is between 0 and 100 per cent – or set by some institutional agreement – for instance, the fluctuation bands of the EMS exchange rates. Second, the definition of the bounds does not only involve the boundaries but also include the deterministic component and the LRV, which are relevant elements to compute unit root tests and perform the statistical inference. Therefore, it would be the case that the bounds change not because of a change in the boundaries, but because of the deterministic component. This setup defines different situations depending on the way that the structural breaks affect the parameters of the model:

- Case A: the boundaries are constant $\left[\underline{b_t}, \overline{b_t}\right] = \left[\underline{b}, \overline{b}\right]$ $\forall t$ and the magnitude of the structural breaks (θ_i) is fixed for all i, $i = 1, 2, \ldots, m$, giving $[\sigma^{-1}T^{-1/2}\left(\underline{b} - d_t\right), \sigma^{-1}T^{-1/2}\left(\overline{b} - d_t\right)]$ $= [\underline{c}, \overline{c}] + o\left(1\right)$ $\forall t$, so that, the presence of structural breaks does not affect the bounds of the limited time series, i.e., $\left[\underline{c_i}, \overline{c_i}\right] = [\underline{c}, \overline{c}]$ $\forall i$, $i = 1, 2, \ldots, m + 1$.

- Case B: the boundaries change but the magnitude of the structural breaks (θ_i) is fixed for all i, $i = 1, 2, \ldots, m$. So, the bounds are defined by:

$$\left[\frac{\left(\underline{b_t} - d_t\right)}{\sigma T^{1/2}}, \frac{\left(\overline{b_t} - d_t\right)}{\sigma T^{1/2}}\right] = \left[\underline{c_i}, \overline{c_i}\right] + o\left(1\right) \text{ for } T_{i-1}^0 < t \leq T_i^0, \tag{5}$$

$i = 1, 2, \ldots, m+1$. In this case, the change of $\overline{c_i}$ between two consecutive regimes is given by $\overline{c_i} - \overline{c_{i-1}} = \sigma^{-1}T^{-1/2}\left(\overline{\gamma_i} - \overline{\gamma_{i-1}}\right) = \overline{\pi_i}$. Similarly, $\underline{c_i} - \underline{c_{i-1}} = \sigma^{-1}T^{-1/2}\left(\underline{\gamma_i} - \underline{\gamma_{i-1}}\right) = \underline{\pi_i}$.

- Case C: the boundaries do not change, but the magnitude of the structural breaks is non-negligible, which can be modeled as a function of T, i.e., in our case:

$$\theta_i = \kappa_i T^{1/2}, \tag{6}$$

$i = 1, 2, \ldots, m$. The specification of the break magnitude given in (6) has also been suggested in Leybourne and Newbold (2000), Kim et al. (2000) and Harvey et al. (2001) as a way to capture structural breaks where the magnitude of the shifts is large. In this

case, the bounds are defined as in (5) and the change of $\overline{c_i}$ between two consecutive regimes is given by $\overline{c_i} - \overline{c_{i-1}} = -\sigma^{-1}T^{-1/2}\theta_i = -\sigma^{-1}\kappa_i$. Note that the same applies for the lower bound, i.e., $\underline{c_i} - \underline{c_{i-1}} = -\sigma^{-1}\kappa_i$.

- Case D: the boundaries change and the magnitude of the structural breaks is non-negligible, $\theta_i = \kappa_i T^{1/2}$, $i = 1, 2, \ldots, m$. In this case the bounds are defined as in (5) and the change of $\overline{c_i}$ between two consecutive regimes is given by $\overline{c_i} - \overline{c_{i-1}} = \sigma^{-1}T^{-1/2}\left(\overline{\gamma_i} - \overline{\gamma_{i-1}}\right) - \sigma^{-1}T^{-1/2}\theta_i = \overline{\pi_i} - \sigma^{-1}\kappa_i$ whereas, for the lower bound, we have $\underline{c_i} - \underline{c_{i-1}} = \underline{\pi_i} - \sigma^{-1}\kappa_i$.

Some remarks are in order. First, in Case A the limits that affect the time series do not change because θ_i is fixed $\forall i$, so that the change in the bounds is $\underline{c_i} - \underline{c_{i-1}} = \overline{c_i} - \overline{c_{i-1}} = o\,(1)$. This does not happen for the other cases. Second, we have shown different ways in which the bounds that affect the time series can change – i.e., Cases B to D – although these specifications lead to the same result as for the unit root testing procedure. However, there is a qualitative difference between the changes driven by the boundaries and those caused by the structural breaks with non-negligible effects. On the one hand, the boundaries are usually known a priori so we do not need to estimate them. Furthermore, in most cases, the boundaries remain constant by the definition of the time series that is analyzed – for instance, the boundaries for the unemployment rate are [0, 100]. On the other hand, the presence of structural breaks with non-negligible effects can relax the assumption of known break points since, in this case, it is possible to get a consistent estimate of the break fractions – further details are given below. When the effect of the structural breaks is negligible, we need to return to the known break points assumption, given that it is not possible to get a consistent estimation of the break fractions. Finally, it is worth noticing that hybrid specifications can be defined where, in some of the regimes, the change in the limits is driven by changes in the boundaries whereas, in other regimes, the change in the limits is caused by the presence of structural breaks with non-negligible effects.

3 Unit root test statistics for broken bounded processes

In this section we present five unit root test statistics that can be applied to test for the presence of a unit root when working with bounded integrated processes with multiple structural breaks. First, we have the three M-type test statistics in Ng and Perron (2001), i.e., the modified Sargan-Bhargava $MSB\left(\lambda^0\right) = \left(T^{-2}\hat{\sigma}^{-2}\sum_{t=1}^{T}\hat{x}_{t-1}^2\right)^{1/2}$, $MZ_\alpha\left(\lambda^0\right) = \left(T^{-1}\hat{x}_T^2 - \hat{\sigma}^2\right)\left(2T^{-2}\sum_{t=1}^{T}\hat{x}_{t-1}^2\right)^{-1}$ and $MZ_t\left(\lambda^0\right) = MZ_\alpha\left(\lambda^0\right) \times MSB\left(\lambda^0\right)$, where $\hat{x}_t$ denotes the OLS detrended variable and $\hat{\sigma}^2$ is a consistent estimate of the LRV. The estimation of the LRV σ^2 can be obtained using either a non-parametric estimator $\left(s_{NP}^2\right)$, as in Phillips and Perron (1988), or a parametric estimator $\left(s_{AR}^2\right)$, as in Ng and Perron (2001). The non-parametric estimator is given by $s_{NP}^2 = T^{-1}\sum_{t=2}^{T}\hat{e}_t^2 + 2T^{-1}\sum_{j=1}^{T-1}k\left(j, M_T\right)\sum_{t=j+1}^{T}\hat{e}_t\hat{e}_{t-j}$, where $\hat{e}_t$ is obtained from the OLS estimation of $\Delta\hat{x}_t = \sum_{i=1}^{m}\delta_i D\left(T_i^0\right)_t + \beta_0\hat{x}_{t-1} + e_t$, where $D\left(T_i^0\right)_t = 1$ for $t = T_i^0 + 1$ and 0 otherwise, and $k\left(j, M_T\right)$ is some kernel that weights the sample autocovariances and M_T is a bandwidth that acts as a truncation lag parameter when $k\left(j, M_T\right) = 0$ for $|j| > M_T$. The parametric estimator is based on the augmented Dickey-Fuller (ADF) regression equation:

$$\Delta\hat{x}_t = \sum_{i=1}^{m}\sum_{j=0}^{k}\delta_{i,j}D\left(T_i^0\right)_{t-j} + \beta_0\hat{x}_{t-1} + \sum_{j=1}^{k}\beta_j\Delta\hat{x}_{t-j} + e_{tk}, \tag{7}$$

so that $s_{AR}^2 = \hat{\sigma}_k^2/(1 - \hat{\beta}(1))^2$, where $\hat{\beta}(1) = \sum_{j=1}^{k}\hat{\beta}_j$ and $\hat{\sigma}_k^2 = (T-k)^{-1}\sum_{t=k+1}^{T}\hat{e}_{tk}^2$ are obtained from the OLS estimation of (7) with the lag length selected as suggested in Perron and Qu (2007).[1] The fourth unit root test that can be computed is the pseudo t-ratio ADF statistic $\left(ADF\left(\lambda^0\right)\right)$ to test the null hypothesis that $\beta_0 = 0$ in (7). Finally, we also consider the Variance Ratio $\left(VR\left(\lambda^0\right)\right)$ statistic in Breitung (2002) $VR\left(\lambda^0\right) = T^{-2}\sum_{t=1}^{T}\left(\sum_{j=1}^{t}\hat{x}_j\right)^2 / \sum_{t=1}^{T}\hat{x}_t^2$. The limit distribution of the unit root test statistics is given in the following Theorem.

---

[1] We are using the OLS detrended version of the M tests proposed by Ng and Perron (2001), since the non-centrality parameter required in the quasi-GLS detrending depends on the bounds – see Carrion-i-Silvestre and Gadea (2013) for the non breaks case. It would be possible to design an iterative estimation procedure to compute the non-centrality parameter, although the goal of this paper is to propose test statistics that allow the presence of structural breaks when testing the unit root hypothesis in limited time series.

Theorem 1 *Let $\{x_t\}_{t=1}^T$ be a stochastic process with its DGP given by (1)-(3). Further, let $T_i^0 \to \infty$ and $T \to \infty$ in such a way that $T_i^0/T \to \lambda_i^0$, $0 < \lambda_i^0 < 1$, where λ_i^0 denotes the break fraction parameter, $i = 1, 2, \ldots, m$. Following Chang and Park (2002), let k in (7) be chosen in a way that $1/k + k^2/T \to 0$ as $T \to \infty$. Under the null hypothesis that $\alpha = 1$ in (3), the $MSB(\lambda^0)$, $MZ_\alpha(\lambda^0)$, $MZ_t(\lambda^0)$, $ADF(\lambda^0)$ and $VR(\lambda^0)$ statistics converge, for Cases B, C and D, to*

$$
MSB\left(\lambda^0\right) \;\Rightarrow\; \left(\sum_{i=1}^{m+1}\left(\lambda_i^0 - \lambda_{i-1}^0\right)^2 \int_0^1 V_{\underline{c_i}}^{\overline{c_i}}\left(s_i\right)^2 ds_i\right)^{1/2}
$$

$$
MZ_\alpha\left(\lambda^0\right) \;\Rightarrow\; \frac{\left(\left(1-\lambda_m^0\right) V_{\underline{c_{m+1}}}^{\overline{c_{m+1}}}\left(1\right)\right)^2 - 1}{2\sum_{i=1}^{m+1}\left(\lambda_i^0 - \lambda_{i-1}^0\right)^2 \int_0^1 V_{\underline{c_i}}^{\overline{c_i}}\left(s_i\right)^2 ds_i}
$$

$$
VR\left(\lambda^0\right) \;\Rightarrow\; \frac{\int_0^1 \left[\sum_{j=1}^{i-1}\left(\lambda_j^0 - \lambda_{j-1}^0\right)\int_0^1 V_{\underline{c_j}}^{\overline{c_j}}\left(s_j\right) ds_j + \left(\lambda_i^0 - \lambda_{i-1}^0\right)\int_{\lambda_{i-1}^0}^r V_{\underline{c_i}}^{\overline{c_i}}\left(s_i\right) ds_i\right]^2 dr}{\sum_{i=1}^{m+1}\left(\lambda_i^0 - \lambda_{i-1}^0\right)^2 \int_0^1 V_{\underline{c_i}}^{\overline{c_i}}\left(s_i\right)^2 ds_i},
$$

where $V_{\underline{c_i}}^{\overline{c_i}}\left(s_i\right) = B_{\underline{c_i}}^{\overline{c_i}}\left(s_i\right) - s_i \int_0^1 B_{\underline{c_i}}^{\overline{c_i}}\left(k_i\right) dk_i$, $i = 1, 2, \ldots, m+1$, are independent demeaned regulated Brownian motions and $\Rightarrow$ denotes weak convergence of the associated measure of probability, with the convention that $\lambda_0^0 = 0$ and $\lambda_{m+1}^0 = 1$. The limiting distribution for the $MZ_t\left(\lambda^0\right)$ test can be obtained from the fact that $MZ_t\left(\lambda^0\right) = MZ_\alpha\left(\lambda^0\right) \times MSB\left(\lambda^0\right)$, a limiting distribution that is equivalent to that of the $ADF\left(\lambda^0\right)$ test statistic.

The proof is given in the appendix. The critical values have been obtained through Monte Carlo simulations for the pair of values $\left[\underline{c_i}, \overline{c_i}\right]$ given by $-\underline{c_i} = \overline{c_i} = \{0.3, 0.5, 0.7, 0.9, 1.1, 1.5\}$ with all possible combinations of break fraction parameters $\lambda = \{0.1, 0.2, \ldots, 0.9\}$ taken in groups of up to $m = 4$ elements – i.e., we allow for up to $m = 4$ structural breaks. We have used 300 steps to approximate the Brownian motion functionals and 10,000 replications.[2,3] It is worth mentioning that the effect of $\left[\underline{c_i}, \overline{c_i}\right]$ on the asymptotic critical values depends on how narrow the limits are set. As expected, the asymptotic critical values do not change signifi-

---

[2] In order to summarize the information, we have estimated response surfaces for the one and two structural breaks cases. Further, it should be stressed that, although we have computed critical values for the symmetric bounds case, our results are valid for general cases where $-\underline{c_i} \neq \overline{c_i}$. A Matlab routine is available from the authors to compute critical values for whatever combination of parameters is needed.

[3] A small set of simulations was carried out increasing the number of steps to 1,000, although the estimated critical values did not change significatively.

cantly when the range of variation is wide, although this is not the case for narrower ranges. It would be also possible to compute (case specific) critical values using bootstrap methods as in Cavaliere and Xu (2014), which is expected to improve finite sample performance of the statistics.

Some remarks are in order. First, the limit distribution of the statistics depends both on the number (m) and position (λ^0) of the structural breaks, although this dependence is not only due to the effect that the level shifts has on the deterministic component of the time series – i.e., the specification given by Case C – but also on the bounds that limit the Brownian motion – i.e., the specification given by Case B. Second, these limiting distributions admit, as a particular case, the situation where the time series is unbounded – i.e., $-\underline{c_i} = \overline{c_i} = \infty \ \forall i$, $i = 1, 2, \ldots, m+1$ – so the regulated Brownian motion becomes a standard Brownian motion. Third, note that the limiting distribution of the $MSB\left(\lambda^0\right)$ statistic has been expressed as a weighted sum of independent functionals of demeaned regulated Brownian motions. Fourth, although we have considered that the bounds are symmetric, it would be possible that the time series is only limited below or above. In this case, the results that have been obtained are still valid, if the unconstrained limit is replaced by either $\underline{c_i} = -\infty$ (bounded above) or $\overline{c_i} = \infty$ (bounded below), depending on the case. Finally, the limit distribution of the test statistics when the structural breaks do not change the limits (Case A) is given in the following Corollary.

Corollary 1 *Let $\{x_t\}_{t=1}^{T}$ be a stochastic process with its DGP given by (1)-(3). Further, let $T_i^0 \to \infty$ and $T \to \infty$ in a such way that $T_i^0/T \to \lambda_i^0$, $0 < \lambda_i^0 < 1$, where λ_i^0 denotes the break fraction parameter, $i = 1, 2, \ldots, m$. Following Chang and Park (2002), let k in (7) be chosen in a such way that $1/k + k^2/T \to 0$ as $T \to \infty$. Under the null hypothesis that $\alpha = 1$ in (3), the $MSB\left(\lambda^0\right)$, $MZ_\alpha\left(\lambda^0\right)$, $MZ_t\left(\lambda^0\right)$, $ADF\left(\lambda^0\right)$ and $VR\left(\lambda^0\right)$ statistics converge for Case A to the limiting distributions in Theorem 1 but imposing $\left[\underline{c_i}, \overline{c_i}\right] = \left[\underline{c}, \overline{c}\right] \ \forall i, \ i = 1, 2, \ldots, m+1$.*

The proof follows from the one for Theorem 1 specifying $\left[\underline{c_i}, \overline{c_i}\right] = \left[\underline{c}, \overline{c}\right] \ \forall i, \ i = 1, 2, \ldots, m+$ 1. Note that, although the presence of structural breaks does not affect the limits of the process, they affect the limiting distribution of the unit root tests, so test statistics for bounded

time series without structural breaks cannot be used in this case. Finally, the limiting distribution of the $MZ_\alpha\left(\lambda^0\right)$, $MZ_t\left(\lambda^0\right)$ and $ADF\left(\lambda^0\right)$ test statistics with $-\underline{c_i} = \overline{c_i} = \infty \ \forall i$, coincides with the ones obtained in Perron (1990), for the one break case, and, implicitly, in Clemente, Montañés and Reyes (1998), for the two breaks case.

4   Unknown breaks case

So far, we have assumed that both the number and the position of the structural breaks are known. However, it is important to design a procedure to estimate the break points when they are unknown. As mentioned above, consistent estimation of the break fractions is only possible when the structural breaks have non-negligible effects on the time series, structural breaks that are modeled as in (6). The approach that we propose in this paper is based on the minimization of the sum of squared residuals under the null hypothesis of unit root ($S(1, \lambda)$), which implies the OLS estimation of the model:

$$\Delta x_t = \sum_{i=1}^{m} \delta_i D\left(T_i\right)_t + e_t, \tag{8}$$

$t = 2, \ldots, T$, over all possible combinations of break points for a given number of structural breaks m. The estimated break dates $(\hat{T}_1, \ldots, \hat{T}_m)'$ are obtained as $\hat{\lambda} = \arg\min_{\lambda \in \Lambda(\epsilon)} S(1, \lambda)$, where the infimum is taken over all possible break fractions defined on the set $\Lambda\left(\epsilon\right) = \left\{(\lambda_1, \lambda_2, \ldots, \lambda_m)'; \ |\lambda_i - \lambda_{i-1}| \geq \epsilon \ (i = 1, 2, \ldots, m-1), \lambda_1 \geq \epsilon, \lambda_m \leq 1 - \epsilon\right\}$, with ϵ being a trimming parameter that dictates the minimal length of a segment – a common value in the related literature is $\epsilon = 0.15$. The following proposition establishes the consistency and the rate of convergence of the estimated break fractions.

Theorem 2 *Let $\{x_t\}_{t=1}^{T}$ be a stochastic process with its DGP given by (1)-(3), with $\alpha = 1$, $m > 0$ and $\theta_i = \kappa_i T^{1/2} \neq 0$, $i = 1, 2, \ldots, m$. Let $\hat{\lambda} = \arg\min_{\lambda \in \Lambda(\epsilon)} S(1, \lambda)$, with $S(1, \lambda)$ being the sum of squared residuals of (8). Then, as $T \to \infty$, $\left\|\hat{\lambda} - \lambda^0\right\| = O_p\left(T^{-1}\right)$ and the $MSB(\hat{\lambda})$, $MZ_\alpha(\hat{\lambda})$, $MZ_t(\hat{\lambda})$, $ADF(\hat{\lambda})$ and $VR(\hat{\lambda})$ statistics converge to the limiting distributions in Theorem 1.*

10

The proof is given in the appendix. In order to estimate m, we could think of defining a maximum number of structural breaks ($m^{\max}$) and select m by minimizing the Bayesian information criterion (BIC) as proposed in Yao (1988) or the modified BIC information criterion in Liu, Wu and Zidek (1997) – henceforth, LWZ information criterion.[4] The following theorem establishes the properties of using these information criteria to estimate m.

Theorem 3 *Let $\{x_t\}_{t=1}^T$ be a stochastic process with its DGP given by (1)-(3), with $\alpha = 1$, $m > 0$ and $\theta_i = \kappa_i T^{1/2} \neq 0$, $i = 1, 2, \ldots, m$, being $m^{\max}$ the maximum number of structural breaks that are allowed. Let us denote by $\hat{m}_{BIC}$ and $\hat{m}_{LWZ}$ the estimated number of structural breaks obtained by minimizing the BIC and LWZ information criteria, respectively. Thus, as $T \rightarrow \infty$, $\hat{m}_{BIC} \overset{p}{\rightarrow} m_{BIC}$ and $\hat{m}_{LWZ} \overset{p}{\rightarrow} m_{LWZ}$, $m \leq \{m_{BIC}, m_{LWZ}\} \leq m^{\max}$, where $\overset{p}{\rightarrow}$ denotes convergence in probability. Further, under the assumption that $e_t \sim iid\ N(0,1)$ in (8), $\hat{m}_{LWZ} \overset{p}{\rightarrow} m$ as $T \rightarrow \infty$.*

The proof is given in the appendix[5]. Theorem 3 states that the use of the BIC and LWZ information criteria might lead, in general, to an over-estimation of m. However, there might be situations in which the LWZ information criteria leads to a consistent estimation of m, which will depend on the extreme value behavior of e_t in (8). Thus, under the restrictive assumption that $e_t \sim iid\ N(0,1)$ in (8), it can be shown that the LWZ information criterion provides a consistent estimation of m, but that the BIC information criterion does not – see the appendix for further details. With this potential risk, it should be borne in mind that an over-estimation of m might imply a power loss of the unit root test statistics. In order to overcome this drawback, we also explore the use of the approach in Burridge and Taylor (2006) – hereafter, BT – who design a procedure for additive outlier detection using extreme value theory. Burridge and Taylor (2006) show that their procedure works reasonably well for different probability distributions of the disturbance term, preventing us from making

---

[4]In principle, it would be possible to design a testing procedure to estimate the number of structural breaks following the approach in Harvey, Leybourne and Taylor (2010). However, this testing procedure will depend on the true number of structural breaks and the values of the bounds, which makes the approach very cumbersome from an implementation point of view. The use of information criteria avoids these complications and, as will be shown by the simulation experiments below, produce satisfactory results in terms of empirical size and power of the unit root test statistics.

[5]We are indebted to an anonymous reviewer for providing insightful help in the proof of this theorem.

assumptions on the probability distribution of e_t in (8).[6] The performance of the informa-tion criteria and the algorithm proposed in Burridge and Taylor (2006) in finite samples are investigated in the Monte Carlo section below.

Once consistent estimates of the break fractions are obtained, the bounds can be computed as in Cavaliere and Xu (2014) from $\underline{\hat{c}_i} = (\underline{b_t} - \hat{d}_t)/(\hat{\sigma}T^{1/2})$ and $\overline{\hat{c}_i} = (\overline{b_t} - \hat{d}_t)/(\hat{\sigma}T^{1/2})$ for $T_{i-1}^0 < t \leq T_i^0$, where it is assumed that the boundaries $[\underline{b_t}, \overline{b_t}]$ are known. In order to compute these bounds, we propose using the deterministic component estimated under the null hypothesis that $\alpha = 1$, which defines $\hat{d}_t = x_0 + \sum_{i=1}^m \hat{\delta}_i DU_{i,t}$, where x_0 is the initial condition, $\hat{\delta}_i$ denotes the OLS parameter estimates of (8) and $\hat{\sigma}$ indicates a consistent estimate of the LRV – either s_{NP}^2 or s_{AR}^2.[7]

5   Finite sample performance

In this section we investigate the performance of the statistics that have been proposed in the paper. The DGP is given by (1) to (3) and (4), where $x_t \in [\underline{b}, \overline{b}]$ and $\varepsilon_t \sim iid\ N(0,1)$. The bounds are set as $[T^{-1/2}(\underline{b} - d_t), T^{-1/2}(\overline{b} - d_t)] = [\underline{c_i}, \overline{c_i}] + o(1)$ for $T_{i-1}^0 < t \leq T_i^0$, $i = 1, 2$, with $T_0^0 = 0$, $T_1^0 = \lambda_1^0 T$ and $T_2^0 = T$. Note that this setup corresponds to Case C above. The Monte Carlo experiment considers the presence of one unknown structural break where the true break point is located at $\lambda_1^0 = 0.5$.[8] The bounds for the first regime are $[\underline{c_1}, \overline{c_1}] = [-0.3, 0.3]$ whereas, for the second regime, they are defined as $[\underline{c_2}, \overline{c_2}] = [\underline{c_1} - \kappa, \overline{c_1} + \kappa]$ with $\kappa = \{0.2, 0.4, 0.6, 0.8\}$. The empirical size is investigated by setting $\alpha = 1$ in (3), whereas for the empirical power we specify $\alpha = 0.8$ in (3). The selection of the order of the model used in the estimation of the parametric LRV is based on the modified Akaike information

---

[6]We have computed the weights that are used in the BT approach taking into account the presence of bounds, although the differences are negligible. Therefore and provided that no gains are obtained from the use of bound-specific weights, we suggest using the weights in Burridge and Taylor (2006).

[7]Note that in practice, we can compute either a parametric or a non-parametric estimation of the bounds, depending on whether we use a parametric or a non-parametric estimation of the LRV. In what follows, we will make abuse of language and talk about the parametric and non-parametric versions of the ADF and VR unit root test statistics, where this distinction will come from the different estimates that can be used to estimate the LRV, which defines the estimate of the bounds.

[8]Simulations are also conducted for other values of λ^0, for the one break case, and have also considered the presence of two structural breaks. Due to space constraints, we only report some of the simulations that have been carried out. The rest are available in the companion appendix.

criterion (AIC) proposed in Ng and Perron (2001) with a maximum of $\left[4\left(T/100\right)^{1/4}\right]$ lags. The non-parametric LRV is estimated with the Newey and West (1994) automatic procedure using the quadratic spectral window. The sample sizes are $T = \{100, 200, 300\}$ and 1,000 replications are used. In all cases, the 5% level of significance asymptotic critical values are used.

5.1 Estimation of the number of structural breaks

Table 1 presents the performance of the BIC, LWZ and BT statistics when used to estimate m following the procedure described in Section 4. The true number of structural breaks is either $m = 0$ or $m = 1$, with the maximum number of m set at $m^{\max} = 2$. As can be seen, the BIC over-estimates m in all cases, tending to select $m^{\max}$. The use of the LWZ statistic shows better performance. When $m = 0$, the selection of the true m improves as T gets larger, reaching values around 0.85 for $T = 300$, regardless of the order of integration of the process. When $m = 1$, the LWZ tends to under-estimate m for $\kappa = 0.2$, but renders good results for the other values of κ. These results confirm the theoretical analysis above. The BT statistic outperforms the BIC and LWZ when $m = 0$, where the probability of selecting the true m is close to one in all cases. In general, the superiority of BT over the BIC and LWZ statistics is also evident when $m = 1$ and $\kappa > 0.4$, although the LWZ statistic outperforms the BT when $\kappa \leq 0.4$ – in these cases, the BT statistic under-estimates m.

In summary, the BIC is liberal in the selection of m, whereas the BT statistic outperforms the other statistics when $m = 0$, but under-estimates m when κ is small. The behavior of the LWZ statistic lies between these extremes. Thus, LWZ presents mild over-estimation problems of m when $m = 0$, but its performance is similar to that of BT when $m = 1$ with large T and κ. Moreover, LWZ outperforms the BT statistic when $m = 1$ with small κ.

5.2 Empirical size and power of the unit root tests

We investigate the performance of the $MSB(\hat{\lambda})$ and $ADF(\hat{\lambda})$ statistics when the BIC, LWZ and BT statistics are used to estimate m, with $m^{\max} = 2$ – note that we are focusing on

the case where both the number and position of the structural breaks are unknown. The results for the other unit root test statistics, which are available in the companion appendix, are similar. Let us first focus on the empirical size. Table 2 shows that, in general, the $MSB(\hat{\lambda})$ statistic presents under-rejection problems, which are more noticeable when m is estimated using the BT approach. Under-rejection is less pronounced when the BIC is used and when the LRV is non-parametrically estimated. The $ADF(\hat{\lambda})$ statistic also shows mild size distortions (under-rejection) although, in most cases, the empirical size is close to the nominal one. Note that the performance of the $ADF(\hat{\lambda})$ statistic does not depend on the estimator that is used to estimate the LRV.

The estimation of m reduces the empirical power of the unit root test statistics when compared to the (unreported) results for known m. As expected, the empirical power of the statistics increases with T and κ, so that, for large values of T and κ, the empirical power is the same as if m is known. If we compare the results based on the BIC, LWZ and BT statistics, we can conclude that, for small values of κ, the BIC-based unit root test statistics are the most powerful. This is a direct consequence of the over-estimation of m that is observed for the BIC – it has been shown that the LWZ and BT statistics tend to under-estimate m when κ is small. As κ increases, the LWZ-based unit root test statistics tend to improve their empirical power, outperforming the BIC-based ones – see, for instance, the results for $\kappa = 0.4$. Finally, when κ is large, the empirical power of both the LWZ and BT-based unit root test statistics is similar.

The method that is used in the estimation of the LRV affects the empirical power of the test statistics. Thus, for a given method of estimating m, we observe that the empirical power of the $MSB(\hat{\lambda})$ statistic is higher when using the non-parametric LRV. This is to be expected, given the size distortion (under-rejection) observed for this statistic when the parametric LRV is used. For the $ADF(\hat{\lambda})$ statistic, the use of the parametric LRV gives higher empirical power.

The simulations that have been conducted in this section reveal that the over-estimation of m leads to unit root tests statistics with both an empirical size that is controlled – in some cases, under-rejection problems are observed – and decent empirical power figures. The

14

under-estimation of m can result in important reductions of the empirical power of the unit root test statistics, a situation that can appear for small values of κ. Therefore, in such cases, it is preferable to over-estimate m.

6   Empirical illustration

There are several studies in the economic literature that assess the properties of the unemployment rate as a way to test whether it behaves in accordance with the (perfect) hysteresis hypothesis or in accordance with the natural unemployment rate theory. In this section, we use the monthly U.S. unemployment rate from January 1948 through November 2014 taking into account the existence of bounds and structural breaks when testing the unit root hypothesis – see Figure 1.[9]

Table 3 reports the M-type, VR and ADF test statistics.[10] When bounds are ignored, the null hypothesis of unit root is rejected by the M-type and ADF test statistics at different levels of significance, regardless of the LRV estimate that is used. However, if we consider the limited nature of the variable, the evidence against the unit root hypothesis is weaker. In order to assess whether this conclusion is robust to the introduction of structural breaks, our analysis has specified up to two structural breaks, with the LWZ selecting one structural break and the BIC and BT statistics selecting two. The fact that the BT statistic detects more structural breaks than the LWZ, together with its tendency to either under (narrow bands) or correctly (wide bands) estimate m, would suggest specifying $m = 2$ – as shown below, wide bands are estimated. However and in order to check the robustness of the conclusions, we proceed considering both $m = 1$ and $m = 2$. The first structural break is placed in December 1974, which reflects the increase of unemployment as a result of the seventies oil crises – Perron (1990) fixed the exogenous break in the fourth quarter of 1973. The second structural break is detected in March 1958, capturing the decrease of unemployment from the expansion during the Kennedy-Johnson administrations until the macroeconomic turmoil of the seventies – see

---

[9]We have used the information provided by the U.S. Bureau of Labor Statistics (BLS): Current Population Survey (CPS), for the seasonally adjusted total unemployment rate.

[10]The parametric long-run variance has been estimated using up to the integer value of $\left[12 \left(T/100 \right)^{1/4} \right]$ lags.

McNown and Seip (2011).

Let us first focus on analysis for $m = 1$. When the parametric LRV estimate is used, the M-test statistics leads to the rejection of the null hypothesis of BI(1) at the 1% level of significance, whereas the VR and ADF tests reject it at the 10% and 5%, respectively. Evidence against the null hypothesis is reduced for the M-test statistics when the non-parametric LRV estimate is used, although the null hypothesis is still rejected at the 10% level of significance. The converse situation is found for the VR and ADF tests, where evidence against the null hypothesis increases. In general, the picture is not modified when we specify $m = 2$. Thus, the same conclusion is reached for the M-type tests. The ADF statistic still rejects the null hypothesis at the 5% level of significance, regardless of the LRV estimate that is used. The exception is the VR statistic, where the null hypothesis is not rejected. Finally, Table 3 also presents the estimated bounds for each specification, which define wide bands. For the parametric LRV-based results, the interval ranks are between 6.4 and 9.5, depending on the model, whereas for the non-parametric LRV-based results, they are between 9.6 and 10. This feature reinforces the specification of $m = 2$, as argued above.

To sum up, evidence against the unit root hypothesis in the postwar U.S. unemployment rate is found when structural breaks are considered, a conclusion that is in accordance with earlier results such as those in Perron (1990). Furthermore, contrary to previous analyses, our results are robust to the presence of theoretical bounds and structural breaks.

7   Conclusions

The paper extends the concept of bounded integrated processes to cover situations where the time series is affected by the presence of multiple structural breaks. Structural breaks enter into the model in two different ways, either affecting the deterministic component or the boundaries component. It has been shown that the limiting distribution of the proposed unit root tests depends not only on the number and position of the structural breaks, but also on the way that the structural breaks affect the (deterministic and boundaries) components of the model. The paper has also shown that consistent estimates of the break fractions are possible

if the magnitude of the level shifts is non-negligible. This allows us to relax the assumption of known breaks in some situations. The Monte Carlo simulations that have been carried out allow us to compare the different unit root tests that have been designed. Simulation results evidence that the M-type tests show a better overall performance. In some cases they are conservative, but they show good empirical power.

The proposal is illustrated analyzing the U.S. unemployment rate. Taking into account both the presence of multiple structural breaks and the limited nature of this time series lead to robust conclusions of the statistical properties of the unemployment rate, which is characterized as a broken bounded I(0) process.

Our approach has assumed that the boundaries are known, something that is not a limitation if we are dealing with variables that are bounded by definition or by construction. However, it would be of interest to consider our framework in cases where it is suspected that the variable is regulated but the boundaries are not known. In this case, Cavaliere and Xu (2014) suggest that a reasonable range of bounds can be inferred from historical observations or from the economic theory. Although the unknown bounds case is outside the scope of this paper, we think that this topic is very interesting from an empirical point of view, so we plan to address it in future research.

A   Appendix

Lemma 1 *Let $\{y_t\}_{t=1}^{T}$ be a stochastic process generated according to (3) and (4) with $\alpha = 1$, and satisfying Assumptions 1 to 3. As $T \to \infty$, $\sigma^{-1}T^{-1/2}y_t \Rightarrow B_{\underline{c}}^{\bar{c}}(r)$, where $B_{\underline{c}}^{\bar{c}}(r)$ denotes a standard regulated Brownian motion, with $\underline{c} \leq 0 \leq \bar{c}$, $\underline{c} \neq \bar{c}$.*

See Theorem 1 in Cavaliere (2005) for the proof.

A.1   Proof of Theorem 1

Note that the deterministic component given in (2) can be written orthogonally as $x_t = \mu_i + y_t$ for $T_{i-1}^0 < t \leq T_i^0$, $i = 1, 2, \ldots, m + 1$, with the convention that $T_0^0 = 0$ and $T_{m+1}^0 = T$, where

the parameter of the deterministic regressor in (2) is a function of the ones in (2). The OLS detrended variable can be obtained as $\hat{x}_t = y_t - z_t\left(\lambda^0\right)\left(z'\left(\lambda^0\right)z\left(\lambda^0\right)\right)^{-1}z'\left(\lambda^0\right)y = [M_{\lambda^0}x]_t = [M_{\lambda^0}y]_t$, where $M_{\lambda^0} = I - P_{\lambda^0} = I - z\left(\lambda^0\right)\left(z'\left(\lambda^0\right)z\left(\lambda^0\right)\right)^{-1}z'\left(\lambda^0\right)$ with $z\left(\lambda^0\right)$ being a $(T\times(m+1))$ matrix with typical rows given by $z_t\left(\lambda^0\right) = [1\left(T_0^0 < t \le T_1^0\right)\ 1\left(T_1^0 < t \le T_2^0\right)$ $\ldots\ 1\left(T_m^0 < t \le T_{m+1}^0\right)]$, where $1\left(\cdot\right)$ denotes the indicator function. The rescaled detrended variable is $T^{-1/2}\hat{x}_t = T^{-1/2}y_t - T^{-1/2}z_t\left(\lambda^0\right)D\left(Dz'\left(\lambda^0\right)z\left(\lambda^0\right)D\right)^{-1}Dz'\left(\lambda^0\right)y$, where $D = diag(T^{-1/2},\ldots,T^{-1/2})$. Then, $T^{-1/2}\hat{x}_t = T^{-1/2}y_t - \frac{r-\lambda_{i-1}^0}{\lambda_i^0-\lambda_{i-1}^0}T^{-3/2}\sum_{t=T_{i-1}+1}^{T_i}y_t$ for $T_{i-1}^0 < t \le T_i^0$, that in the limit converges to $T^{-1/2}\hat{x}_t \Rightarrow \sigma[B_{\underline{c_i}}^{\overline{c_i}}\left(r\right) - \frac{r-\lambda_{i-1}^0}{\lambda_i^0-\lambda_{i-1}^0}\int_{\lambda_{i-1}}^{\lambda_i}B_{\underline{c_i}}^{\overline{c_i}}\left(s\right)ds]$ for $T_{i-1}^0 < t \le T_i^0$. The computation of the MSB statistic can be expressed as $MSB\left(\lambda^0\right) = (T^{-2}\hat{\sigma}^{-2}(\sum_{t=1}^{T_1^0}\hat{x}_t^2$
$+\sum_{t=T_1^0+1}^{T_2^0}\hat{x}_t^2+\cdots+\sum_{t=T_m^0+1}^{T-1}\hat{x}_t^2))^{1/2} \Rightarrow (\sum_{i=1}^{m+1}\int_{\lambda_{i-1}^0}^{\lambda_i^0}(B_{\underline{c_i}}^{\overline{c_i}}\left(r\right) - \frac{r-\lambda_{i-1}^0}{\lambda_i^0-\lambda_{i-1}^0}\int_{\lambda_{i-1}}^{\lambda_i}B_{\underline{c_i}}^{\overline{c_i}}\left(s\right)ds)^2dr)^{1/2}$, provided that $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$ – see below. Note that we can define $s_i = \left(r - \lambda_{i-1}^0\right)/\left(\lambda_i^0 - \lambda_{i-1}^0\right)$, $s_i \in [0,1]$, and $\int_{\lambda_{i-1}^0}^{\lambda_i^0}(B_{\underline{c_i}}^{\overline{c_i}}\left(r\right) - \frac{r-\lambda_{i-1}^0}{\lambda_i^0-\lambda_{i-1}^0}\int_{\lambda_{i-1}}^{\lambda_i}B_{\underline{c_i}}^{\overline{c_i}}\left(s\right)ds)^2dr = \left(\lambda_i^0 - \lambda_{i-1}^0\right)^2\int_0^1(B_{\underline{c_i}}^{\overline{c_i}}\left(s_i\right) - b_i\int_0^1 B_{\underline{c_i}}^{\overline{c_i}}\left(k_i\right)dk_i)^2ds_i$ so that the limit distribution of the $MSB\left(\lambda^0\right)$ test can be expressed as a weighted sum of independent functionals of demeaned regulated Brownian motions, i.e., $MSB\left(\lambda^0\right)$ $\Rightarrow \left(\sum_{i=1}^{m+1}\left(\lambda_i^0 - \lambda_{i-1}^0\right)^2\int_0^1 V_{\underline{c_i}}^{\overline{c_i}}\left(s_i\right)^2ds_i\right)^{1/2}$, where $V_{\underline{c_i}}^{\overline{c_i}}\left(s_i\right) = B_{\underline{c_i}}^{\overline{c_i}}\left(s_i\right) - s_i\int_0^1 B_{\underline{c_i}}^{\overline{c_i}}\left(k_i\right)dk_i$ denotes a demeaned regulated Brownian motion, and $k_i = \left(s - \lambda_{i-1}^0\right)/\left(\lambda_i^0 - \lambda_{i-1}^0\right)$, $k_i \in [0,1]$. Let us now focus on the $MZ_\alpha\left(\lambda^0\right)$ test $MZ_\alpha\left(\lambda^0\right) = \left(T^{-1}\hat{x}_T^2 - \hat{\sigma}^2\right)\left(2T^{-2}\sum_{t=1}^T\hat{x}_{t-1}^2\right)^{-1} \Rightarrow (((1-\lambda_m^0)V_{\underline{c_{m+1}}}^{\overline{c_{m+1}}}\left(1\right))^2 - 1)(2\sum_{i=1}^{m+1}\left(\lambda_i^0 - \lambda_{i-1}^0\right)^2\int_0^1 V_{\underline{c_i}}^{\overline{c_i}}\left(s_i\right)^2ds_i)^{-1}$. The limit distribution of the $MZ_t\left(\lambda^0\right) = MZ_\alpha\left(\lambda^0\right)\times MSB\left(\lambda^0\right)$ statistic follows from the derivations above. For the $ADF\left(\lambda^0\right)$ statistic, note that the ADF regression equation in (7) can be expressed as:

$$\Delta\hat{x}_t = \beta_0\hat{x}_{t-1} + \hat{\xi}_t'\varphi + e_{tk}, \tag{9}$$

with $\hat{\xi}_t = (D\left(T_1^0\right)_t, D\left(T_1^0\right)_{t-1},\ldots,D\left(T_1^0\right)_{t-k},\ldots,D\left(T_m^0\right)_t, D\left(T_m^0\right)_{t-1},\ldots,D\left(T_m^0\right)_{t-k}, \Delta\hat{x}_{t-1},\ldots,\Delta\hat{x}_{t-k})'$, $\varphi = (\delta_1,\ldots,\delta_m,\beta_1,\ldots,\beta_k)'$, $\delta_i = (\delta_{i,0},\delta_{i,1},\ldots,\delta_{i,k})'$, $i = 1,\ldots,m$, and $\beta\left(1\right) = \sum_{j=1}^k\beta_j$. Following Chang and Park (2002), we can define $A_T\left(\lambda\right) = \sum_{t=1}^T\hat{x}_{t-1}e_{tk} - \left(\sum_{t=1}^T\hat{x}_{t-1}\hat{\xi}_t'\right)\left(\sum_{t=1}^T\hat{\xi}_t\hat{\xi}_t'\right)^{-1}\left(\sum_{t=1}^T\hat{\xi}_t e_{tk}\right)$, $B_T\left(\lambda\right) = \sum_{t=1}^T\hat{x}_{t-1}^2 - \left(\sum_{t=1}^T\hat{x}_{t-1}\hat{\xi}_t'\right)\left(\sum_{t=1}^T\hat{\xi}_t\hat{\xi}_t'\right)^{-1}\left(\sum_{t=1}^T\hat{\xi}_t\hat{x}_{t-1}\right)$ and $C_T\left(\lambda\right) = \sum_{t=1}^T e_{tk}^2 - \left(\sum_{t=1}^T e_{tk}\hat{\xi}_t'\right)\left(\sum_{t=1}^T\hat{\xi}_t\hat{\xi}_t'\right)^{-1}\left(\sum_{t=1}^T\hat{\xi}_t e_{tk}\right)$, so that $\hat{\beta}_0 = B_T^{-1}\left(\lambda\right)A_T\left(\lambda\right)$, with the variance of the error term given by $\hat{\sigma}_k^2 = T^{-1}(C_T\left(\lambda\right) - A_T^2\left(\lambda\right)B_T^{-1}\left(\lambda\right))$ and the

variance of the estimated β_0 parameter given by $s_T^2\left(\hat{\beta}_0\right) = \hat{\sigma}_k^2 B_T^{-1}(\lambda)$. Using these elements and following Chang and Park (2002), it is straightforward to see that as $T \to \infty$, $T^{-1}A_T(\lambda)$ $\Rightarrow \frac{1}{2}\sigma^2\left[\left(\left(1-\lambda_m^0\right) V_{\underline{c_{m+1}}}^{\overline{c_{m+1}}}(1)\right)^2 - 1\right]$, $T^{-2}B_T(\lambda) \Rightarrow \sigma^2 \sum_{i=1}^{m+1}\left(\lambda_i^0 - \lambda_{i-1}^0\right)^2 \int_0^1 V_{\underline{c_i}}^{\overline{c_i}}(s_i)^2 ds_i$ and $T^{-1}C_T(\lambda) \xrightarrow{p} \sigma_k^2$. Note that, as shown in Chang and Park (2002) and Cavaliere and Xu (2014), $\hat{\beta}(1) \xrightarrow{p} \beta(1)$ and $C(1) = 1/(1 - \beta(1))$ so that $\hat{\sigma}^2 = \hat{\sigma}_k^2/(1-\hat{\beta}(1))^2 \xrightarrow{p} \sigma^2$, where $\hat{\sigma}^2$ is defined here using the parametric version of the LRV estimator $\left(s_{AR}^2\right)$ – the same result can be established using the non-parametric version of the LRV estimator. Then, the $ADF\left(\lambda^0\right)$ statistic is computed as $ADF\left(\lambda^0\right) = \hat{\sigma}^{-1}B_T^{-1/2}(\lambda)A_T(\lambda)$, which in the limit converges to

$$ADF\left(\lambda^0\right) \Rightarrow \frac{\left(\left(1-\lambda_m^0\right) V_{\underline{c_{m+1}}}^{\overline{c_{m+1}}}(1)\right)^2 - 1}{2\left(\sum_{i=1}^{m+1}\left(\lambda_i^0 - \lambda_{i-1}^0\right)^2 \int_0^1 V_{\underline{c_i}}^{\overline{c_i}}(s_i)^2 ds_i\right)^{1/2}}, \tag{10}$$

which coincides with the limiting distribution of the $MZ_t\left(\lambda^0\right)$ derived above.

Finally, the numerator of the $VR\left(\lambda^0\right)$ statistic for a generic $T_{i-1} < t \leq T_i$ involves the square term of $T^{-3/2}\sum_{j=1}^t \hat{x}_j \Rightarrow \sum_{j=1}^{i-1}\sigma\int_{\lambda_{j-1}^0}^{\lambda_j^0}\left(B_{\underline{c_j}}^{\overline{c_j}}(s)\,ds - \frac{s-\lambda_{j-1}^0}{\lambda_j^0-\lambda_{j-1}^0}\int_{\lambda_{j-1}^0}^{\lambda_j^0} B_{\underline{c_j}}^{\overline{c_j}}(b)\,db\right)ds$ $+ \sigma\int_{\lambda_{i-1}^0}^r\left(B_{\underline{c_i}}^{\overline{c_i}}(s)\,ds - \frac{s-\lambda_{i-1}^0}{\lambda_i^0-\lambda_{i-1}^0}\int_{\lambda_{i-1}^0}^{\lambda_i^0} B_{\underline{c_i}}^{\overline{c_i}}(b)\,db\right)ds$, or, equivalently, $T^{-3/2}\sum_{j=1}^t \hat{x}_j \Rightarrow \sigma[\sum_{j=1}^{i-1}$ $\left(\lambda_j^0 - \lambda_{j-1}^0\right)\int_0^1 V_{\underline{c_j}}^{\overline{c_j}}(s_j)\,ds_j + \left(\lambda_i^0 - \lambda_{i-1}^0\right)\int_{\lambda_{i-1}^0}^r V_{\underline{c_i}}^{\overline{c_i}}(s_i)\,ds_i]$, so that the limit distribution of the $VR\left(\lambda^0\right)$ statistic can be expressed as

$$VR\left(\lambda^0\right) \Rightarrow \frac{\int_0^1\left[\sum_{j=1}^{i-1}\left(\lambda_j^0 - \lambda_{j-1}^0\right)\int_0^1 V_{\underline{c_j}}^{\overline{c_j}}(s_j)\,ds_j + \left(\lambda_i^0 - \lambda_{i-1}^0\right)\int_{\lambda_{i-1}^0}^r V_{\underline{c_i}}^{\overline{c_i}}(s_i)\,ds_i\right]^2 dr}{\sum_{i=1}^{m+1}\left(\lambda_i^0 - \lambda_{i-1}^0\right)^2 \int_0^1 V_{\underline{c_i}}^{\overline{c_i}}(s_i)^2 ds_i}.$$

A.2 Proof of Theorem 2

Let λ^0 to denote the true break fraction and λ a generic one. Note that the sum of squared residuals of the model in first differences, $S(1,\lambda)$, can be computed as $S(1,\lambda) = Q^1(\lambda)$ $- 2G^1(\lambda) + u^{1\prime}M_\lambda^1 u^1$, where $d^1(\lambda) = z^1(\lambda) - z^1\left(\lambda^0\right)$, $M_\lambda^1 = I - P_\lambda^1$, and $P_\lambda^1 = z^1(\lambda)$ $\left(z^{1\prime}(\lambda) z^1(\lambda)\right)^{-1} z^{1\prime}(\lambda)$, while the superscript 1 denotes the first difference of the corresponding element. When the break fraction is correctly specified, $\lambda^0 = \lambda$, and $S\left(1,\lambda^0\right) = u^{1\prime}M_{\lambda^0}^1 u^1$. The difference between these two sums of squared residuals is given by $S(1,\lambda) - S\left(1,\lambda^0\right)$

19

$= Q^1(\lambda) - 2G^1(\lambda) + u^{1\prime}M_\lambda^1 u^1 - u^{1\prime}M_{\lambda^0}^1 u^1$. Consider first $Q^1(\lambda) = \psi'd^{1\prime}(\lambda)\,d^1(\lambda)\,\psi -$

$\psi'd^{1\prime}(\lambda)\,P_\lambda^1 d^1(\lambda)\,\psi$. For the sake of simplicity, assume that there is only one structural

break, although the derivations are valid for the multiple structural break. Then, the first

element of $Q^1(\lambda)$ is $\psi'd^{1\prime}(\lambda)\,d^1(\lambda)\,\psi = 2\theta^2$, while the second element is $\psi'd^{1\prime}(\lambda)\,P_\lambda^1 d^1(\lambda)\,\psi$

$= \psi'd^{1\prime}(\lambda)\,z^1(\lambda)\left(z^{1\prime}(\lambda)\,z^1(\lambda)\right)^{-1}z^{1\prime}(\lambda)\,d^1(\lambda)\,\psi = \theta^2$, so that $Q^1(\lambda) = 2\theta^2 - \theta^2 = \kappa^2 T\,\tilde{I}$,

with $\tilde{I} = 1\left(\left|T_1 - T_1^0\right| \neq 0\right)$. Further, note that $T^{-1}Q^1(\lambda) > 0$. Consider now $G^1(\lambda) =$

$u^{1\prime}d^1(\lambda)\,\psi - u^{1\prime}z^1(\lambda)\,z^{1\prime}(\lambda)\,d^1(\lambda)\,\psi = -u_{T_1^0+1}^1 \theta\,\tilde{I}$, so that $T^{-1}G^1(\lambda) = o_p(1)$. Finally, we

have $u^{1\prime}M_\lambda^1 u^1 - u^{1\prime}M_{\lambda^0}^1 u^1 = -u^{1\prime}P_\lambda^1 u^1 + u^{1\prime}P_{\lambda^0}^1 u^1 = -\left(u_{T_1+1}^1\right)^2 + \left(u_{T_1^0+1}^1\right)^2 \leq O_p(1)\,\tilde{I}$. Col-

lecting terms, $T^{-1}\left(S(1,\lambda) - S\left(1,\lambda^0\right)\right) = T^{-1}Q^1(\lambda) + o_p(1) = \kappa^2\tilde{I} + o_p(1)$. Furthermore,

note that for a given break fraction, λ, it is always satisfied that $T^{-1}S(1,\lambda) \leq T^{-1}S\left(1,\lambda^0\right)$.

Let us now suppose that $\lambda \overset{p}{\nrightarrow} \lambda^0$. According to this inequality, we will need that, for large

T, $T^{-1}Q^1(\lambda) \leq 0$, but we have shown that $Q^1(\lambda) > 0$ when $\lambda \neq \lambda^0$. Then, a contradiction

appears, and the only way that the inequality is satisfied is when $\lambda \overset{p}{\to} \lambda^0$. The, minimization

of the $S(1,\lambda)$ over all possible values of the break date results in consistent estimate of the

break date. As mentioned above, although derivations have considered only one structural

break, they are also valid for the multiple break case.

To establish the convergence rate of the break fraction parameter, we first define the sets

$V_\epsilon = \left\{T_k : \left|T_k - T_1^0\right| < \epsilon T\right\}$ for $\epsilon \in (0,1)$ and $V_\epsilon(C) = \left\{T_k : \left|T_k - T_1^0\right| < \epsilon T, \left|T_k - T_1^0\right| > C\right\}$ for $C > 0$, so $V_\epsilon(C) \subset V_\epsilon$. Note that $S(1,\lambda) \leq S\left(1,\lambda^0\right)$ with probability 1 and

$\Pr(T_1 \in V_\epsilon) \to 1$, as $T \to \infty$. Given previous results, there is a constant $C > 0$ such that

$\Pr(\min_{\lambda T \in V_\epsilon(C)} \frac{S(1,\lambda) - S(1,\lambda^0)}{|\lambda - \lambda^0|T} \leq 0) < \zeta$ for some small $\zeta > 0$ because, when C is properly cho-

sen, $\Pr(\min_{\lambda T \in V_\epsilon(C)} \frac{S(1,\lambda) - S(1,\lambda^0)}{|\lambda - \lambda^0|T} \leq 0) = \Pr(\min_{\lambda T \in V_\epsilon(C)} \frac{Q^1(\lambda) - 2G^1(\lambda) + u^{1\prime}M_\lambda^1 u^1 - u^{1\prime}M_{\lambda^0}^1 u^1}{|\lambda - \lambda^0|T} \leq 0)$

$< \zeta$, because $\frac{Q^1(\lambda) - 2G^1(\lambda) + u^{1\prime}M_\lambda^1 u^1 - u^{1\prime}M_{\lambda^0}^1 u^1}{|\lambda - \lambda^0|T} = O(1) - 2O_p(T^{-1/2}) + \frac{O_p(1)\,\tilde{I}}{C}$, with the $O(1)$ and

$O_p(1)$ terms positive. Hence, $\frac{S(1,\lambda) - S(1,\lambda^0)}{|\lambda - \lambda^0|T} > 0$ on $V_\epsilon(C)$ with large probability. This implies

that the minimum cannot be achieved on $V_\epsilon(C)$ and, thus, $\Pr\left(T|\hat{\lambda} - \lambda^0| \geq C\right) \leq \zeta$, so that

$\left(\hat{\lambda} - \lambda^0\right) = O_p\left(T^{-1}\right)$.

In order to show that the limiting distribution of the unit root test statistics using $\hat{\lambda}$

are the same as the ones using λ^0 we can write the OLS estimated parameters as $\hat{\mu} = z(\hat{\lambda})$

$$\left(z'(\hat{\lambda})z(\hat{\lambda})\right)^{-1} z'(\hat{\lambda})x = z(\hat{\lambda}) \left(z'(\hat{\lambda})z(\hat{\lambda})\right)^{-1} z'(\hat{\lambda})z(\lambda^0)\mu + z(\hat{\lambda}) \left(z'(\hat{\lambda})z(\hat{\lambda})\right)^{-1} z'(\hat{\lambda})y = \mu + z(\hat{\lambda})$$

$\left(z'(\hat{\lambda})z(\hat{\lambda})\right)^{-1} z'(\hat{\lambda})y + z(\hat{\lambda}) \left(z'(\hat{\lambda})z(\hat{\lambda})\right)^{-1} z'(\hat{\lambda}) \left[z(\lambda^0) - z(\hat{\lambda})\right] \mu$. Using these elements, we can compute the detrended variable that is used in the computation of all statistics as

$$
\begin{aligned}
\hat{x} &= x - z(\hat{\lambda})\hat{\mu} = z(\lambda^0)(\mu - \hat{\mu}) + y + \left[z(\lambda^0) - z(\hat{\lambda})\right]\hat{\mu} \\
&= y - z(\lambda^0)\left(z'(\hat{\lambda})z(\hat{\lambda})\right)^{-1} z'(\hat{\lambda})y \\
&\quad - z(\lambda^0)\left(z'(\hat{\lambda})z(\hat{\lambda})\right)^{-1} z'(\hat{\lambda})\left[z(\lambda^0) - z(\hat{\lambda})\right]\mu + \left[z(\lambda^0) - z(\hat{\lambda})\right]\hat{\mu}.
\end{aligned}
$$

Note that the two first terms are $\left\| y - z(\lambda^0)\left(z'(\hat{\lambda})z(\hat{\lambda})\right)^{-1} z'(\hat{\lambda})y \right\| = O_p(T)$. The third term,

$\left\| z(\lambda^0)\left(z'(\hat{\lambda})z(\hat{\lambda})\right)^{-1} z'(\hat{\lambda})\left[z(\lambda^0) - z(\hat{\lambda})\right]\mu \right\| = \mu'\left[z(\lambda^0) - z(\hat{\lambda})\right]' z(\hat{\lambda})\left(z'(\hat{\lambda})z(\hat{\lambda})\right)^{-1} z'(\lambda^0)z(\lambda^0)$

$\left(z'(\hat{\lambda})z(\hat{\lambda})\right)^{-1} z'(\hat{\lambda})\left[z(\lambda^0) - z(\hat{\lambda})\right]\mu)^{1/2} = O_p(T^{1/2})$ and the fourth term, $\left\| \left[z(\lambda^0) - z(\hat{\lambda})\right]\hat{\mu}\right\| =$

$(\hat{\mu}'\left[z(\lambda^0) - z(\hat{\lambda})\right]' \left[z(\lambda^0) - z(\hat{\lambda})\right]\hat{\mu})^{1/2} = \left(\left|\hat{\lambda} - \lambda^0\right| O_p(T^2)\right)^{1/2} = O_p(T^{1/2})$. Therefore, we can see that the last two terms are asymptotically negligible compared to the first two so that, in the limit, the test statistics using the estimated break dates converge to the limiting distribution of the test statistics where the break dates are known.

A.3 Proof of Theorem 3

The $S(1, \lambda)$ of the OLS estimated model defined in (8) can be written as $S(1, \lambda) = \Delta x_t^2 - \sum_{i=1}^{m} \Delta x_{T_i+1}^2 = \Delta x_t^2 - \sum_{i=1}^{m} \theta_i^2 + O_p(1)$. The BIC and LWZ information criteria that we propose to use in the paper are given by $BIC = \log(S(1,\lambda)/(T-1-m)) + m\log(T-1-m)/(T-1-m)$ and $LWZ = \log(S(1,\lambda)/(T-1-m)) + 0.299m\log(T-1-m)^{2.1}/(T-1-m)$. Let us focus on the BIC. As can be seen, the first element of the right hand side is $\log(\frac{1}{T-1-m}(\Delta x_t^2 - \sum_{i=1}^{m} \theta_i^2 + O_p(1))) = \log(\frac{1}{T-1-m}(\Delta x_t^2 - \sum_{i=1}^{m} \kappa_i^2 T + O_p(1)))$ provided that the magnitude of the structural breaks is non-negligible, i.e., $\theta_i = \kappa_i T^{1/2}$. As $T \to \infty$, this element converges to $\log(\sigma_{\Delta x}^2 - \sum_{i=1}^{m} \kappa_i^2)$, where $\sigma_{\Delta x}^2$ denotes the variance of Δx_t, whereas the second element of the right hand side converges to zero for fixed m. Therefore, accounting for the presence of structural breaks reduces the sum of the squared residuals in the limit, leading to the detection of the structural breaks when using the BIC information criterion – the same applies

to the LWZ information criterion.

However, this does not necessarily imply a consistent estimation of the number of structural breaks, i.e., it might be the case that the use of information criteria leads to over-detection of the number of structural breaks. In order to investigate this issue, let us consider the case where there is one structural break, $m = 1$, and we compute the BIC information criterion for the models that consider one (BIC_1) and two structural breaks (BIC_2). Let us denote by $S(1, \lambda_1)$ and $S(1, \lambda_2)$ the sum of squared residuals of the model (8) for the one and two breaks cases, respectively. It can be seen that $BIC_2 - BIC_1$
$\approx \ln\left(T^{-1}S(1, \lambda_2)\right) + 2T^{-1}\ln(T) - \ln\left(T^{-1}S(1, \lambda_1)\right) - T^{-1}\ln(T) = \ln\left(1 + \frac{S(1, \lambda_2) - S(1, \lambda_1)}{S(1, \lambda_1)}\right)$
$+ T^{-1}\ln(T)$, where $S(1, \lambda_2) - S(1, \lambda_1) = -\Delta x_{T_2+1}^2$, with T_2 denoting the second break date. We need $BIC_2 - BIC_1 > 0$ if the correct number of structural breaks is going to be selected. The order of $S(1, \lambda_2) - S(1, \lambda_1)$ is determined by the extreme value of e_t in (8), which depends on the underlying assumptions that we make about its distribution. In the simplest case where $e_t \sim iidN(0, 1)$ in (8), $\max_t |e_t| = O_p(\sqrt{\ln(T)})$ so that $BIC_2 - BIC_1 \approx -O_p(T^{-1}\ln(T)) + T^{-1}\ln(T)$, which is not guaranteed to be positive in the limit. Therefore and even in this simple case, the BIC information criterion might lead to an over-specification of the number of structural breaks, i.e., $\hat{m}_{BIC} \geq m$. This situation is not found when using the LWZ information criterion, since $LWZ_2 - LWZ_1 \approx -O_p(T^{-1}\ln(T)) + 0.299 T^{-1}\ln(T)^{2.1}$, which is positive in the limit, although this result relies on the assumption that is made about the extreme behavior of e_t in (8).

Finally, it is worth noticing that we have based the proof of this theorem on the assumption that the structural breaks have non-negligible effects. If the magnitude of the structural breaks is fixed – i.e., asymptotically negligible – these information criteria do not necessarily lead to a consistent estimate of the number of structural breaks.

References

[1] Breitung J. (2002) Nonparametric tests for unit roots and cointegration. *Journal of Econometrics* 108, 343-363.

[2] Burridge, P., Taylor, A.M.R. (2006) Additive outlier detection via extreme-value theory. *Journal of Time Series Analysis* 27, 685-701.

[3] Carrion-i-Silvestre J.L., Gadea M.D. (2013) GLS-based unit root tests for bounded processes. *Economics Letters* 120, 184-187.

[4] Cavaliere G. (2005) Limited time series with a unit root. *Econometric Theory* 21, 907-945.

[5] Cavaliere G., Xu F. (2014) Testing for unit roots in bounded nonstationary time series. *Journal of Econometrics* 178, 259-272.

[6] Chang Y., Park G.Y. (2002) On the asymptotics of ADF tests for unit roots. *Econometric Reviews* 21, 431-447.

[7] Clemente J., Montañés A., Reyes M. (1998) Testing for a unit root in variables with a double change in the mean. *Economics Letters* 59, 175-182.

[8] Harvey D.I., Leybourne S.J., Newbold P. (2001) Innovational outlier unit root tests with an endogeneously determined break in level. *Oxford Bulletin of Economics and Statistics* 63, 559-575.

[9] Harvey D.I., Leybourne S.J., Taylor A.M.R. (2010) Robust methods for detecting multiple level breaks in autocorrelated time series. *Journal of Econometrics* 157, 342-358.

[10] Kim T., Leybourne S.J., Newbold P. (2000) Spurious rejections by Perron tests in the presence of a break. *Oxford Bulletin of Economics and Statistics* 62, 433-444.

[11] Leybourne S.J., Newbold P. (2000) Behavior of the standard and symmetric Dickey-Fuller-type tests when there is a break under the null hypothesis. *Econometrics Journal* 3, 1-15.

[12] Liu J., Wu S., Zidek J.V. (1997) On Segmented Multivariate Regressions. *Statistica Sinica* 7, 497-525.

[13] McNown R., Seip K.L. (2011) Periods and structural breaks in US economic history 1959–2007. *Journal of Policy Modeling* 33, 169-182..

[14] Newey W.K., West K.D. (1994) Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61, 631-653.

[15] Ng S., Perron P. (2001) Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519-1554.

[16] Perron P. (1989) The great crash, the oil price shock and the unit root hypothesis. *Econometrica* 57, 1361-1401.

[17] Perron P. (1990) Testing for a unit root in a time series with a changing mean. *Journal of Business and Economic Statistics* 8, 153-162.

[18] Perron P., Qu Z. (2007) A simple modification to improve the finite sample properties of Ng and Perron's unit root tests. *Economics Letters* 94, 12-19.

[19] Phillips P.C.B., Perron P. (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335-346.

[20] Yao Y.C. (1988) Estimating the number of change-points via Schwarz' criterion. *Statistics and Probability Letters* 6, 181-189.

Table 1: Estimation of the number of structural braks using the BIC, LWZ and BT statistics

			BIC						LWZ						BT					
			I(1)			I(0)			I(1)			I(0)			I(1)			I(0)		
m	$\bar{c}$	$\hat{m}\backslash T$	100	200	300	100	200	300	100	200	300	100	200	300	100	200	300	100	200	300
0	0.3	0	0.034	0.018	0.005	0.045	0.025	0.014	0.508	0.713	0.811	0.565	0.778	0.862	0.990	0.995	0.989	0.989	0.997	0.993
		1	0.237	0.167	0.129	0.231	0.188	0.150	0.369	0.253	0.168	0.328	0.197	0.132	0.009	0.005	0.009	0.009	0.003	0.006
		2	0.729	0.815	0.866	0.724	0.786	0.836	0.123	0.034	0.021	0.107	0.026	0.007	0.002	0.001	0.002	0.002	0.001	0.001
0	0.5	0	0.041	0.018	0.010	0.053	0.026	0.014	0.544	0.744	0.852	0.581	0.777	0.861	0.992	0.996	0.989	0.992	0.996	0.993
		1	0.226	0.198	0.126	0.243	0.187	0.149	0.339	0.216	0.132	0.320	0.199	0.133	0.007	0.003	0.009	0.007	0.003	0.006
		2	0.733	0.784	0.864	0.705	0.787	0.837	0.117	0.040	0.016	0.099	0.024	0.007	0.001	0.001	0.002	0.001	0.001	0.001
0	0.7	0	0.048	0.022	0.010	0.054	0.026	0.014	0.567	0.751	0.848	0.582	0.777	0.861	0.991	0.996	0.991	0.992	0.996	0.993
		1	0.249	0.188	0.135	0.242	0.187	0.149	0.326	0.217	0.141	0.321	0.199	0.133	0.007	0.003	0.008	0.007	0.003	0.006
		2	0.703	0.790	0.855	0.705	0.787	0.837	0.107	0.032	0.011	0.097	0.024	0.007	0.001	0.001	0.001	0.001	0.001	0.001

			BIC						LWZ						BT					
			I(1)			I(0)			I(1)			I(0)			I(1)			I(0)		
m	κ	$\hat{m}\backslash T$	100	200	300	100	200	300	100	200	300	100	200	300	100	200	300	100	200	300
1	0.2	0	0.020	0.010	0.000	0.030	0.020	0.000	0.300	0.350	0.340	0.460	0.560	0.480	0.820	0.720	0.640	0.990	0.980	0.940
		1	0.200	0.140	0.090	0.220	0.140	0.100	0.470	0.550	0.590	0.400	0.370	0.460	0.180	0.280	0.350	0.010	0.010	0.060
		2	0.780	0.850	0.910	0.750	0.850	0.890	0.230	0.100	0.070	0.150	0.070	0.060	0.000	0.000	0.000	0.000	0.000	0.000
1	0.4	0	0.010	0.000	0.000	0.000	0.000	0.000	0.130	0.120	0.120	0.080	0.010	0.000	0.630	0.430	0.230	0.870	0.580	0.150
		1	0.190	0.140	0.090	0.190	0.120	0.100	0.600	0.730	0.800	0.620	0.850	0.900	0.360	0.570	0.770	0.120	0.420	0.840
		2	0.800	0.860	0.910	0.810	0.880	0.900	0.270	0.150	0.080	0.300	0.140	0.100	0.010	0.000	0.000	0.000	0.000	0.000
1	0.6	0	0.010	0.000	0.000	0.000	0.000	0.000	0.080	0.090	0.080	0.000	0.000	0.000	0.380	0.210	0.150	0.410	0.040	0.000
		1	0.190	0.130	0.090	0.190	0.120	0.100	0.620	0.770	0.840	0.690	0.860	0.900	0.620	0.790	0.840	0.590	0.960	0.990
		2	0.800	0.870	0.910	0.810	0.880	0.900	0.300	0.140	0.080	0.310	0.150	0.100	0.010	0.000	0.010	0.010	0.000	0.000
1	0.8	0	0.010	0.000	0.000	0.000	0.000	0.000	0.070	0.070	0.070	0.000	0.000	0.000	0.260	0.190	0.140	0.060	0.000	0.000
		1	0.190	0.130	0.080	0.190	0.120	0.100	0.650	0.790	0.840	0.690	0.860	0.900	0.730	0.800	0.860	0.930	1.000	1.000
		2	0.800	0.870	0.920	0.810	0.880	0.900	0.280	0.150	0.090	0.310	0.150	0.100	0.010	0.000	0.000	0.010	0.000	0.000

Table 2: Empirical size and power for the unit root test statistics with one unknown structural break and $\lambda^0 = 0.5$. m is selected using the BIC, LWZ and BT statistics

| | | | Empirical size | | | | | | Empirical power | | | | | |
| | | | Parametric | | | Non-parametric | | | Parametric | | | Non-parametric | | |
		κ	100	200	300	100	200	300	100	200	300	100	200	300
BIC	$MSB(\hat\lambda)$	0.2	0.036	0.027	0.026	0.045	0.041	0.036	0.078	0.340	0.799	0.105	0.471	0.881
		0.4	0.024	0.031	0.039	0.027	0.044	0.053	0.036	0.504	0.929	0.060	0.651	0.978
		0.6	0.026	0.037	0.033	0.031	0.047	0.049	0.052	0.559	0.943	0.079	0.691	0.980
		0.8	0.020	0.025	0.021	0.019	0.037	0.032	0.068	0.594	0.940	0.099	0.748	0.988
	$ADF(\hat\lambda)$	0.2	0.069	0.051	0.048	0.064	0.047	0.043	0.156	0.449	0.763	0.147	0.423	0.731
		0.4	0.066	0.050	0.064	0.064	0.042	0.058	0.212	0.649	0.915	0.208	0.614	0.881
		0.6	0.057	0.054	0.054	0.051	0.049	0.052	0.240	0.689	0.932	0.230	0.654	0.902
		0.8	0.045	0.034	0.038	0.044	0.031	0.038	0.244	0.702	0.925	0.235	0.689	0.918
LWZ	$MSB(\hat\lambda)$	0.2	0.024	0.017	0.021	0.030	0.026	0.027	0.026	0.216	0.472	0.037	0.311	0.542
		0.4	0.016	0.015	0.024	0.018	0.030	0.039	0.035	0.677	0.981	0.060	0.826	0.998
		0.6	0.017	0.012	0.017	0.019	0.025	0.034	0.071	0.746	0.990	0.105	0.879	0.996
		0.8	0.009	0.009	0.020	0.011	0.020	0.028	0.100	0.797	0.988	0.150	0.912	0.997
	$ADF(\hat\lambda)$	0.2	0.051	0.036	0.036	0.052	0.032	0.033	0.110	0.316	0.566	0.095	0.250	0.459
		0.4	0.063	0.037	0.052	0.056	0.036	0.049	0.211	0.690	0.963	0.200	0.657	0.935
		0.6	0.050	0.039	0.036	0.046	0.036	0.038	0.266	0.760	0.982	0.251	0.710	0.952
		0.8	0.050	0.029	0.033	0.048	0.028	0.033	0.300	0.780	0.985	0.283	0.753	0.970
BT	$MSB(\hat\lambda)$	0.2	0.006	0.013	0.017	0.011	0.025	0.029	0.003	0.023	0.068	0.012	0.089	0.212
		0.4	0.008	0.015	0.021	0.015	0.028	0.035	0.007	0.314	0.835	0.012	0.371	0.849
		0.6	0.005	0.011	0.015	0.011	0.023	0.025	0.047	0.778	0.992	0.076	0.893	1.000
		0.8	0.004	0.009	0.011	0.008	0.017	0.019	0.106	0.856	0.991	0.163	0.960	1.000
	$ADF(\hat\lambda)$	0.2	0.033	0.032	0.034	0.026	0.027	0.030	0.074	0.156	0.281	0.039	0.065	0.111
		0.4	0.040	0.041	0.045	0.034	0.037	0.042	0.037	0.304	0.821	0.031	0.289	0.796
		0.6	0.039	0.040	0.040	0.037	0.037	0.039	0.173	0.745	0.980	0.165	0.707	0.954
		0.8	0.037	0.033	0.034	0.035	0.032	0.034	0.315	0.805	0.984	0.300	0.770	0.969

Table 3: Unit root tests for the U.S. unemployment rate

LRV	H_0	m	MSB	MZ_α	MZ_t	VR	ADF	$\hat{T}$	$[\underline{c_1}, \overline{c_1}]$	$[\underline{c_2}, \overline{c_2}]$	$[\underline{c_3}, \overline{c_3}]$
P	$I(1)$	0	0.16^a	-20.28^b	-3.18^b	0.018	-3.85^a				
	$BI(1)$	0	0.16^c	-20.28^c	-3.18^c	0.018	-3.85^b		$[-0.3, 9.2]$		
	$BI(1)$	1	0.10^a	-55.26^a	-5.25^a	0.004^b	-4.52^b	74/12	$[-0.2, 6.5]$	$[-0.3, 6.4]$	
	$BI(1)$	2	0.09^a	-61.24^a	-5.53^a	0.004	-4.29^b	(58/3, 74/12)	$[-0.2, 6.2]$	$[-0.3, 6.2]$	$[-0.3, 6.1]$
NP	$I(1)$	0	0.16^b	-19.62^b	-3.13^b	0.018	-3.85^a				
	$BI(1)$	0	0.16^c	-19.62^c	-3.13^c	0.018	-3.85^b		$[-0.3, 9.3]$		
	$BI(1)$	1	0.14^c	-25.28^c	-3.55^c	0.004^b	-4.52^a	74/12	$[-0.3, 9.5]$	$[-0.4, 9.4]$	
	$BI(1)$	2	0.14^c	-25.43^c	-3.56^c	0.004	-4.29^b	(58/3, 74/12)	$[-0.3, 9.7]$	$[-0.4, 9.6]$	$[-0.5, 9.5]$

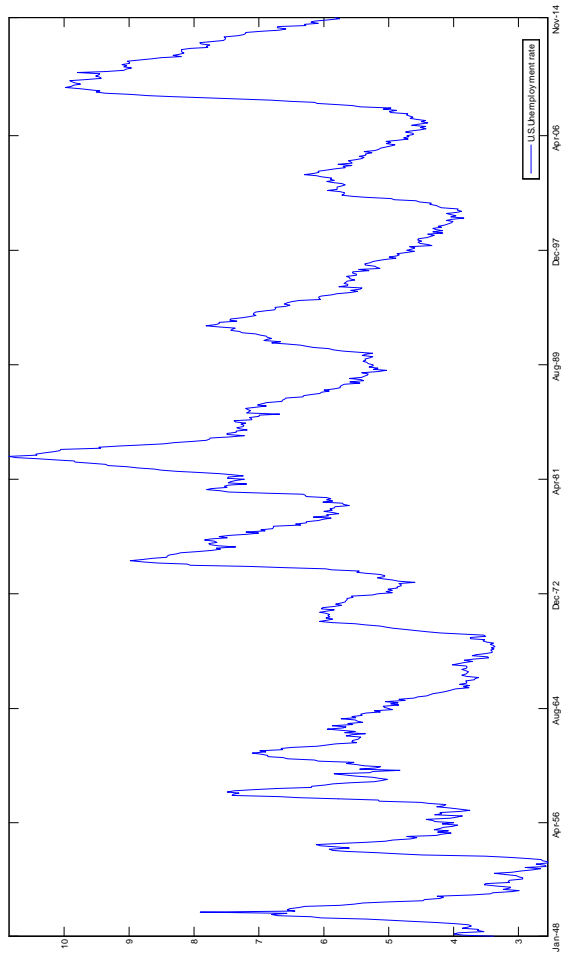Note: P denotes parametric and NP non-parametric. a, b and c indicate rejection of the null hypothesis at 1, 5 and 10%, respectively

Figure 1: U.S. unemployment rate

27