Computational Economics (2025) 65:3111–3159
https://doi.org/10.1007/s10614-024-10651-z

Comparison of the Performance of Structural Break Tests in Stationary and Nonstationary Series: A New Bootstrap Algorithm

Özge Çamalan[1] · Esra Hasdemir[2] · Tolga Omay[1] · Mustafa Can Küçüker[1]

Accepted: 5 June 2024 / Published online: 8 July 2024

Abstract
Structural breaks are considered as permanent changes in the series mainly because of shocks, policy changes, and global crises. Hence, making estimations by ignoring the presence of structural breaks may cause the biased parameter value. In this context, it is vital to identify the presence of the structural breaks and the break dates in the series to prevent misleading results. Accordingly, the first aim of this study is to compare the performance of unit root with structural break tests allowing a single break and multiple structural breaks. For this purpose, firstly, a Monte Carlo simulation study has been conducted through using a generated homoscedastic and stationary series in different sample sizes to evaluate the performances of these tests. As a result of the simulation study, Zivot and Andrews (J Bus Econ Stat 20(1):25–44, 1992) are the best-performing tests in capturing a single break. The most powerful tests for the multiple break setting are those developed by Kapetanios (J Time Ser Anal 26(1):123–133, 2005) and Perron (Palgrave Handb Econom 1:278–352, 2006). A new Bootstrap algorithm has been proposed along with the study's primary aim. This newly proposed Bootstrap algorithm calculates the optimal number of statistically significant structural breaks under more general assumptions. Therefore, it guarantees finding an accurate number of optimal breaks in real-world data. In the empirical part, structural breaks in the real interest rate data of the US and Australia resulting from policy changes have been examined. The results concluded that the bootstrap sequential break test is the best-performing approach due to the general assumption made to cover real-world data.

Keywords Unit root with structural breaks · Monte Carlo simulation · Real interest rate · Bootstrap algorithm

JEL Classification C40 · C53 · C22

Extended author information available on the last page of the article

1 Introduction

In time series studies, the issue of how a series changes over time have significant implications. The tendency of a time series and a given sample period may not always be constant over time, but changes may be due to permanent or temporary shocks. Covariance stationarity is weakened by this issue.

Changes in a time series may be caused by structural breaks. These breaks may arise due to policy changes, global shocks, structural reforms, etc., and generate permanent changes in the series. In other words, when structural breaks in the series are ignored, the estimators of a regression model give biased and misleading results. Thus, it is essential to determine the presence of structural breaks in the series. This phenomenon was first discussed by Perron (1989, 1990) in the literature. Perron (1989) suggested that if there is a structural break in a stationary series and the break date is correctly determined, the null hypothesis of the unit root may be rejected. When a structural break occurs in the series, but the break date is incorrectly specified, the unit root null hypothesis would be rejected based on the incorrect methodology. In this sense, such tests do not have statistical power in testing a unit root. This conclusion was further supported by the power analyses of Perron's (1989, 1990) studies on a small sample.

The stationarity feature of the time series is also crucial for long-term forecasting of the series. Additionally, running a regression model with a nonstationary series may lead to spurious regression and invalidate the classical theory of distribution. In addition to the nonstationarity problem, it is crucial to identify the sources of nonstationarity, which may result from a unit root or structural breaks. That is, determining the presence of structural breaks and the dates of those breaks should be more critical than the unit root problem. This is because structural breaks prevent proper unit root testing by masking stationarity. Hence, unit root tests that allow both structural break and unit root simultaneously are frequently seen in the literature.

Nevertheless, not many structural break unit root tests developed under the stationarity assumption. Therefore, it is vital to differentiate between unit root and structural break as the sources of nonstationarity. Hence, we investigated which test is more powerful in determining the actual break date. For this purpose, random data is generated under the homoscedasticity and stationarity assumption. The Monte Carlo simulation study uses this data to compare the unit root tests with a structural break.

The main contribution of this study is that it is the first study in the literature where a Monte Carlo simulation is conducted to evaluate the statistical performances of the unit root with structural break(s) tests applied in this study. The relevant tests are classified into unit root with a single structural break and unit root with multiple structural break tests. We employed the unit root test with a single structural break developed by Zivot and Andrews (1992) (ZA hereafter), Banerjee et al. (1992) (BLS hereafter), Andrews and Ploberger (1994) (AP hereafter) and Perron (1997), and the unit root test with multiple structural breaks developed by Lumsdaine and Papell (1997) (LP hereafter), Bai and Perron (1998, 2003a) (BP hereafter), Lee and Strazicich (2003) (LS hereafter), Kapetanios (2005) and Perron (2006).

The findings of the Monte Carlo simulation states that the best performed test in determining a single structural break is the ZA test. Kapetanios (2005) and Perron (2006) tests are more powerful ones, and they have equal strength in capturing structural breaks in mean in a series.

In this study, we develop a newly proposed bootstrap methodology along with the main aim of our study. In order to determine the structural break and their break dates more accurately, we proposed a new robust bootstrap method encompassing the null hypothesis, and considering heteroscedasticity and nonstationarity under the null. According to the results of this newly suggested bootstrap method, Kapetanios (2005) is the best performed test in sequential break test. In the empirical part, structural breaks in the real interest rate data of the US and Australia resulting from policy changes have been examined. The results concluded that the bootstrap sequential break test is the best-performing approach due to the general assumption made to cover real-world data.

This paper is organized as follows: Sect. 2 provides a review of the limited literature focusing on the different unit root with structural break tests applied to determine the structural break in mean and explain the reasons of the mean break coming from the economic theory. The third section introduces theoretical foundations of structural break tests applied in this study. In the third section, the Monte Carlo study is explained. In the fourth section, the theoretical foundation of the newly proposed method is exhibited, and empirical research conducted with this newly proposed method is examined in detail. The fifth section presents conclusion.

2 Literature Review

The real interest rate plays a significant role in various economic frameworks, such as monetary transmission, consumption-based asset pricing models, and the neoclassical growth model (Neely and Rapach, 2008). Consequently, considerable research effort has been dedicated to exploring the relationship between the stochastic characteristics of real interest rates as a time series and corresponding policy changes. However, a limited amount of research focuses specifically on the persistence of real interest rates and their underlying determinants.

Many studies in the relevant literature examine the structural changes in the mean of the U.S. real interest rate. Garcia and Perron (1996) analyzed the behavior of the U.S. real interest rate time series from 1961 to 1986, finding that shifts in the mean were mainly influenced by periods of budget deficits and oil shocks rather than changes in policy regimes. Rapach and Wohar (2005) suggest that the structural break in the mean interest rate of the U.S., as observed in 13 other industrialized countries, is attributable to a mean shift in the inflation rate. Caporale and Grier (2005) examine the mean shifts in real interest rates of the U.S. and U.K. They conclude the structural break in the mean is caused by changing the Federal Reserve Chair of the U.S.

Research analyzing the structural breaks in the Australian real interest rate is relatively limited. Felmingham and Mansfield's (2003) study examines the stationarity and structural breaks of short-term and long-term interest rates using the Z.A. and

Nunes et al. (1997) tests. They discover that long-term indicators are trend stationary at the level. In contrast, short-term indicators exhibit stationarity with a structural break. They highlight that the data-generating process of Australia's real interest rate is influenced by political factors, monetary policies, and international effects on Australia's economy, which was regarded as a small economy in 2003. The studies by Mishra et al. (2023) focused on whether shifts in political and bureaucratic regimes resulted in notable changes in Australia's real interest rate. Employing the B.P. structural break test, they identified three distinct structural breaks in the real interest rate in 1980: Q1, 1992: Q2, and 2008: Q4. The study also presents results from the ZA, LP, and Kapetanios (2005) tests. The other research in the literature examines the related issue for the Chinese real interest rate series. Hong et al. (2020) shed light on the factors leading to structural breaks in mean interest rates. Analyzing Chinese data, they used the Bai and Perron (1998, 2003a, 2003b) test to identify structural breaks in various interest rate means. They compared these break dates with interest rate liberalization dates. Their findings revealed structural breaks in mean interest rates over the studied period (Hong et al., 2020).

3 Econometric Methodology

Nelson and Plosser (1982) investigated the stationarity of 14 macroeconomic time series of the US between the years 1907–1970 by employing the ADF unit root test. As a result, they failed to reject the unit root null of 13 series. They concluded that shocks are permanent because stochastic series have a unit root. Following Nelson and Plosser (1982), Perron (1989) suggested that the macroeconomic time series may not only contain unit root, rather there may be temporary fluctuations in the series. Based on this suggestion, Perron (1989) developed the first unit root test with structural break where the break date is determined exogenously. Perron (1989) stated that the null hypothesis of unit root should not be rejected without considering the structural break. Therefore, Perron (1989) added dummy variables to the model to present the break date, which was known in advanced. In addition to that, Perron (1989) allowed structural breaks in the level, slope, or both of the trend functions.

Upon the seminal work of Perron (1989), the related literature has started to grow. Various unit root tests with structural break(s) are developed since then. In this study, the ZA, BLS, Perron (1997) unit root tests are applied to capture a single structural break. The LP, BP, LS, Kapetanios (2005) and Perron (2006) unit root tests are applied to capture multiple structural breaks in the series.

3.1 Unit Root Tests Allowing a Single Structural Break

Following the study of Perron (1989, 1990), Zivot and Andrews (1992) developed a novel unit root with a structural break test focused on determining the structural break date. On the contrary to Perron's (1989, 1990) unit root test, which determined the break date exogenously, Zivot and Andrews (1992) suggested a unit root with structural break where the break date is determined endogenously.

In the ZA test, the null hypothesis stating the unit root is tested against the alternative one stating the trend stationary process. Testing process of the ZA test, an ADF-type test, begins with three models, Model A, Model B, and Model C, which were set up following Perron (1989, 1990). Model A captures a break in the intercept. Model B captures a break in the slope of the trend function, while Model C captures a break both in the intercept and slope. Minimum t-statistic is used for hypothesis testing of three models. The minimum value of the t-statistic is defined as $t_{\hat{\alpha}^i}\left[\hat{\lambda}^i_{\text{inf}}\right] = \inf_{\lambda\in\Lambda} t_{\hat{\alpha}^i}(\lambda), i = A, B, C$ (Zivot & Andrews, 1992).

The second unit root test allowing a single endogenous structural break is developed by Banerjee et al. (1992). The BLS test, an ADF-type test, follows the studies of Rappoport and Reichlin (1989) and Perron (1989, 1990). The assumption of the break date of the BLS test is that it cannot be known a priori. They provide an asymptotic distribution for recursive, rolling, and sequential test statistics. These three test statistics are used in hypothesis testing for the varying unit root/trend cases. The initial model for obtaining the recursive and rolling test statistics is provided in Eq. (1) (Banerjee et al., 1992).

$$\text{Model 1}: \ y_t = \mu_0 + \mu_1 t + \alpha y_{t-1} + \beta(L)\Delta y_{t-1} + \varepsilon_t \quad t = 1, \ldots, t \tag{1}$$

In Eq. (1), $\beta(L)$ is a lag polynomial of the roots of $1 - \beta(L)$ out of the unit circle, and under the unit root null, $\alpha = 1,\ \mu_1 = 0$. ε_t is a martingale difference sequence series.

Recursive test statistics are obtained by using the sub- samples $k = k_0, ..., T$ and $t = 1, ..., k_0$, where T is the sample size and k_0 is the starting value. Rolling statistics are obtained by using sub-samples having a constant fraction and rolling them through the sample (Banerjee et al., 1992).

Sequential test statistics are calculated based on using the full sample subsequently. The model for obtaining sequential test statistics is given in Eq. (2) (Banerjee et al., 1992):

$$\text{Model 2}: y_t = \mu_0 + \mu_1\tau_{1t}(k) + \mu_2 t + \alpha y_{t-1} + \beta(L)\Delta y_{t-1} + \omega' x_{t-1}(k) + \varepsilon_t \quad t = 1, \cdots, T \tag{2}$$

where $\beta(L)$ is a lag polynomial of order p which is already known. In Model 2, there is an additional m-vector of regressors that is given by $x_{t-1}(k)$ which are stationary with a constant zero mean. The term $\tau_{1t}(k)$ shows the possible shift or jump in the trend component at period k. In this framework, Banerjee et al. (1992) considers the two cases given in Eqs. (3) and (4), by following Perron (1989, 1990).

$$\text{Case A } \tau^1_{1t}(\text{k}) = (\text{t} - \text{k})1(\text{t} > \text{k}) \tag{3}$$

$$\text{Case B } \tau^1_{1t}(\text{k}) = 1(\text{t} > \text{k}) \tag{4}$$

The selection of the break date for sequential test statistics is done by the Quandt likelihood ratio (LR) test. Thus, one needs to compute the LR test statistics for at least one break and then choose the maximum value.

The third unit root with structural break test for a single endogeneous structural break is developed by Andrews and Ploberberger (1994). The AP test, an LM-type test, analyzes the null hypothesis of no structural break against the alternative of a structural break. The AP test utilizes exponential Lagrange Multiplier (Exp-LM), exponential Wald and exponential likelihood ratios (LR). At the α significance level, Exp-LM often has the largest weighted mean when compared to the others. Thus, AP uses the Exp-LM test statistics for the testing procedure.

The last structural break test of this group was developed by Perron (1997). Perron focused on cases where the break date was unknown. Perron (1997) states that a finite sample and its distribution depend on the correlation between data and break dates. Perron (1997) considers the sequential test that makes the unit root test values minimum based on the break date. According to Perron (1997), lag selection is a crucial factor in determining break dates.

3.2 Unit Root Tests Allowing Multiple Structural Breaks

After applying unit root tests with a single structural break, the LP, BP, LS, Kapetanios (2005) and Perron (2006) tests are applied to the series to determine whether it exhibits multiple structural breaks.

Lumsdaine and Papell (1997) suggested that conducting a unit root test with a single structural break would provide ambiguous results. Based on their suggestion, they developed a unit root test with two endogenous structural breaks. The LP test, which is an extension of the ADF-type tests, is developed by following the studies of Zivot and Andrews (1992) and Banerjee, Lumsdaine ve Stock (1997). Three models in the LP, Model AA, CA, and CC, are used in the LP. Model AA allows two structural breaks in the mean only. Model CA allows the first structural break in the mean and slope, and the second structural break can only occur in the slope. Model CC does permit two structural breaks both in the mean and slope.

The hypothesis testing procedure of the LP test is based on testing the null hypothesis of the unit root with no structural break against the alternative hypothesis of trend stationarity with two structural breaks in the trend function at two different points. Rejecting the null hypothesis does not necessarily mean rejecting a unit root, but the unit root without breaks. Similarly, the alternative hypothesis does not naturally imply trend stationarity with breaks but may imply a unit root with breaks (Lumsdaine & Papell, 1997).

Lee and Strazicich (2003) estimated two structural break Lagrange Multiplier (LM) unit root test statistics. They followed the testing procedure suggested by Schmidt and Phillips (1992) and Lumsdaine and Papell (1997) to estimate the test statistics.

$$\text{Model AA: } \Delta y_t = \mu + \alpha y_{t-1} + \beta t + \Phi_1 DU_{1,t} + \omega DU_{2,t} + \sum_{j=1}^{k} c_i \Delta y_{t-j} + \varepsilon_t \quad (5)$$

$$\text{Model CA: } \Delta y_t = \mu + \alpha y_{t-1} + \beta t + \Phi_1 DU_{1,t} + \gamma DT_1, t + \omega DU_{2,t} + \sum_{j=1}^{k} c_i \Delta y_{t-j} + \varepsilon_t \quad (6)$$

$$\text{Model CC: } \Delta y_t = \mu + \alpha y_{t-1} + \beta t + \Phi_1 DU_{1,t} + \gamma DT_{1,t} + \omega DU_{2,t} + \psi DT_{2,t} + \sum_{j=1}^{k} c_i \Delta y_{t-j} + \varepsilon_t \quad (7)$$

where $t = 1, \cdots, T$. DU_1 and DU_2 are dummy variables presenting the break dates (TB_1 and TB_2) in mean, respectively. Similarly, DT_1 and DT_2 are dummy variables added to express the structural breaks in trend occurred at TB_1 and TB_2.

$$DU_{1,t} = \begin{cases} 1 & if\ t > TB_1 \\ 0, & otherwise \end{cases} \quad DT_{1,t} = \begin{cases} t - TB_1 & if\ t > TB_1 \\ 0, & otherwise \end{cases}$$
$$DU_{2,t} = \begin{cases} 1 & if\ t > TB_2 \\ 0, & otherwise \end{cases} \quad DT_{2,t} = \begin{cases} t - TB_2 & if\ t > TB_2 \\ 0, & otherwise \end{cases}$$

Optimal lag is chosen by using a general to specific methodology. Break dates are determined by t- statistics that gives the minimum α value.

Lee and Strazicich (2003) generalized the testing procedure suggested by Schmidt and Phillips (1992). The LS test allows for two endogeneous structural breaks under the null and alternative hypotheses. The LS test utilizes the unobserved component model given in Eq. (8).

$$y_t = \delta' Z_t + e_t \quad e_t = \beta e_{t-1} + \varepsilon_t \quad \varepsilon_t \sim iidN\left(0, \sigma^2\right) \quad (8)$$

where Z_t is a vector of exogeneous variables.

The LS unit root test is computed by the model given in Eq. (9).

$$\Delta y_t = \delta' \Delta Z_t + \Phi_1 \tilde{S}_{t-1} + u_t \quad (9)$$

where $\tilde{S}_t = y_t - \tilde{\varphi}_X - Z_t \sigma_X - \tilde{\delta}$, $t = 2, \cdots, T$ and δ' are the coefficients of the estimated regression, $\tilde{\varphi}_X$ is given by $y_1 - Z_1 S'$ where y_1 and Z_1 represents the first observation of y_t and Z_t, respectively. The null and alternative hypotheses are $H_0 : \beta - 1 = \Phi = 0$ and $H_1 : \beta - 1 = \Phi < 0$ are subject to the hypothesis testing by using the LM statistics for $\Phi = 0$. The LS technique employs a grid search procedure to identify the break dates, which are not known. These break dates are determined based on the minimum value of the LM statistic.

$$LM_\tau = \inf_{\hat{\lambda}} \hat{\tau}\left(\hat{\lambda}\right), \ \lambda = (\lambda_1 = TB_1/T, \lambda_2 = TB_2/T) \quad (10)$$

Bai and Perron (1998) unit root test is another unit root test with multiple structural breaks. By applying the BP test, they can examine the presence, number, and position of structural breaks in the series. The BP test analyzes the null of a unit root with breaks against the alternative hypothesis of$l$ number of structural breaks (Bai & Perron, 2003b). Structural breaks are determined sequentially in the BP test. Besides, this test allows heteroscedasticity, serial correlation, and different distributions of error terms.

Kapetanios (2005) and Perron (2006) unit root with multiple structural breaks tests are the other two unit root tests utilized in this study.

The null hypothesis of Kapetanios (2005) test contains unit root null where the alternative hypothesis states stationarity with m- number of structural breaks in constant and/or trend. The undefined number of structural breaks, m, must exceed two but it must be less than or equal to m.

Kapetanios (2005) developed the unit root test based on several existing unit root tests. Kapetanios employs the sequential Dickey–Fuller (DF) t-statistics, following Banerjee et al. (1992) and Zivot and Andrews (1992) for a single break. The testing procedure is started with the following model:

$$y_t = \mu_0 + \mu_1 t + \theta y_{t-1} + \sum_{i=1}^{k} \partial_i \Delta y_{t-1} + \sum_{i=1}^{k} \Phi_i DU_{i,t} + \sum_{i=1}^{k} \alpha_i DT_{i,t} + \varepsilon_t \tag{11}$$

All of $1 - \partial(L)$'s circle are outside of the unit circle. The vector $(\Delta y_{t-1}, ..., \Delta y_{t-k})$ shows the covariance matrix's probability limit. The dummy variables for intercept and trend are $DU_{i,t}$ and $DT_{i,t}$, respectively. These two dummies are defined as $DU_{it} = 1(t > T_{bi}), DT_{it} = 1(t > T_{bi})(t - T_{bi})$. Here, T_{bi} the break date of ith structural break, indicator function is represented as 1(.).

In order to clear up the analysis and identify the vector of regressors, Kapetanios relied on Banerjee et al. (1992) and Lumsdaine and Papell (1997) in addition to Eq. (11). They resolve the Ordinary Least Square (OLS) estimators used in Eq. (11) by specifying the scaling matrix. After applying OLS to the one break case, Kapetanios (2005) applies the grid search implemented by Lumsdaine and Papell (1997) to extend the analysis to m structural breaks. As this method requires a lot of mathematical operations, they follow the grid research employed by Bai and Perron (1998).

A six-step testing structure is proposed by Kapetanios (2005) based on the research of Bai and Perron (1998) and the ordinary least square (OLS) approach as follows: (i) finding one break is the first step in the process of calculating the maximum number of structural breaks. The t-statistics for each potential sample component are then gathered., (ii) the minimum sum of square residuals (SSR) are obtained to determine the break date., (iii) the next break date is obtained by using the break test once more., (iv) similar to the second stage, the next break date is defined where the minimum SSR is obtained., (v) repeat steps 3 and 4 until the minimum t statistics is achieved and, (vi) minimum t statistics is obtained, and m structural break is identified.

In this study, the Perron (2006) test is the last unit root test with multiple structural breaks test applied in this study.[1] Perron asserted that because both structural break and unit root testing methodologies employ the same tools, new research on structural break literature was created along with unit root analysis (Perron, 2006). Consequently, these two work together. Given this fact, Perron (2006) designed a unit root test with

[1] Apart from all this, the Narayan and Popp (2010) test, which is the newest in the literature, allows two breaks. Since the ADF type test they developed is similar to the Lee and Strazicich (2003) test in large samples in terms of power and size, this test was not considered in the study.

Table 1 Summary table of test statistics to estimate the break dates

	Test	Test statistics
Single structural break tests	ZA	Infimum t
	BLS	Quant Likelihood Ratio
	AP	SupLM, SupLR, SupW
	Perron (1997)	Minimum t
Multiple structural break tests	LP	Infimum t
	BP	Argmin SSR
	LS	Infimum t
	Kapetanios (2005)	SSR minimum
	Perron (2006)	Argmin SSR

an undefined break date and concentrated on the decomposition of structural break and unit root.

Perron (2006) applies the sequential test utilized by Bai and Perron (1998) to ascertain the break date by locating the global maximum of the objective function. Perron (2006) uses the ordinary least squares (OLS) method that assigns equal weights to error terms. Breaks can be captured only when variance changes match the breaks in the coefficients. If necessary, different weights should be assigned while taking the variance changes into account. Therefore, the quasi-likelihood method should be considered in Perron (2006) test.

The assumption of a constant trend is incorrect in the case of an unknown break date. Thus, before discussing trend changes, researchers should initially neglect the series' stationarity. In other words, Perron applies this assumption as a diagnostic check. Perron recommends using DF test statistics when there is a constant trend. Yet, when there is a change in trend, the unit root test should be used based on discussing the break date when the SSR is obtained from the relevant regression. When there is only a change in the intercept, the estimate of break date is conflicting under the null. In this case, two statistics can be evaluated together (Perron, 2006).

Perron (2006) discussed that sequential tests usually outperform the other tests. He did, however, issue a caution that by using this method, fewer structural break dates are chosen than actually could be achieved. Perron (2006) also suggested that the number of actual break points may be more than estimated ones because of the problems with the power of the test. Thus, the double maximum test should be firstly used in determining the existence of structural breaks (Perron, 2006).

The abovementioned structural break tests benefit from several test statistics to determine the break dates. These test statistics are given as a summary in Table 1.

4 Monte Carlo Simulation

The estimates of the break date for each test may vary with each trial. In this regard, how closely the estimated values match the actual values is essential. Mean square error (MSE), root moan square error (RMSE), and mean absolute error (MAE) are metrics commonly used as an evaluation indicator. Hansen (2001) utilized MSE in their study as an evaluation indicator, which indicates the residual variance. Another indicator, RMSE, is used to measure the distance between the residuals' estimated and actual values. This makes it easier to find the model's weak points and offers suggestions for future improvements. Additionally, RMSE's ability to draw attention to significant mistakes and outliers. It makes it easier to understand how the model behaves in these situations. Since simulation results inevitably differ from actual values, RMSE plays a crucial role in determining acceptable error boundaries. As a result, it helps choose an appropriate model or simulation result within those constraints. Hence, in this study, the RMSE, which is regarded as the residuals' standard deviation, is used to evaluate the performances of structural break tests. The formula for the RMSE is given below:

$$RMSE = \frac{1}{T}\sum_{j=1}^{T} e_j^2 \tag{12}$$

The RMSE takes any value in the range of 0 to ∞. If the RMSE number is 0, there is no error, and the test performs successfully. A high RMSE value indicates poor performance. Additionally, when penalizing major errors, RMSE is superior to other criteria.

For the Monte Carlo simulation study, 1000 simulation trials have been carried out with different sample sizes as T = 50, T = 100, T = 250, and T = 500. Following that, the actual break dates have been compared with the estimated break dates obtained in each trial of the simulation study.

The break fractions are assigned at the samples' beginning, middle, and end. The simulation procedure is started with models that permit one and two breaks in the mean, as given in Eqs. (13) and (14).

Model 1: Model for a single break in mean:

$$y_t = \mu_1 + \theta_1 Dum1_t + \alpha x_t + e_t \tag{13}$$

Model 2: Model for two breaks in the mean:

$$y_t = \mu_1 + \theta_1 Dum1_t + \theta_2 Dum2_t + \alpha x_t + e_t \tag{14}$$

For each simulation trials, independent and identically distributed random numbers, $X_t \sim N(0, 1)$ were generated by the random number generation. The initial value of y_t is set to 0.5, $\mu_1 = 1$ and $\alpha = 0.5$. $Dum1_t$ and $Dum2_t$ are dummy variables denoting structural breaks. θ_1 is the break magnitude coefficient of the first break and θ_2 is the magnitude of the second break.

In Eq. (13), θ_1 is given different values from 1 to 10. In Eq. (14), θ_1 is given different values between 1 and 5, and θ_2 different values from 1 to 8. λ_1 denotes the first break fraction. The values 0.2, 0.5 and 0.8 are assigned to λ_1 at the beginning, in the middle and at the end of the series, respectively. If $t > T\lambda_1$, $Dum1_t = 1$, otherwise $Dum1_t = 0$. This break fraction definition is constant for both models.

The second break fraction, λ_2, is gradually shifted farther from the first break date. Different values of λ_2 depending on the location of the first break are given below:

If $\lambda_1 = 0.2$ then 0.3, 0.5, 0.7 and 0.9, respectively.

If $\lambda_1 = 0.5$ then 0.7 and 0.9, respectively.

If $\lambda_1 = 0.8$, it takes the value 0.9.

Similarly, if $t > T\lambda_2$, then, $Dum2_t = 1$, and if $t < T\lambda_2$, then $Dum2_t = 0$. By using the randomly generated data, 1000 simulation trials with various sample sizes, break magnitudes, and break fractions for each of the structural break tests are conducted. Based on the simulation study, 1000 structural breaks and break dates were estimated for Model 1 and Model 2.

The findings of the simulation study for one break case are illustrated in Fig. 1 in grid search graphs. As it is seen in Fig. 1, a single break test for the T = 50 does not depend on the break fraction. There is good performance in the Lee and Strazicich (2003), Kapetanios (2005) and Perron (2006) tests, regardless of where the break fraction occurs. Other tests are sensitive to the location of the break fraction and the break magnitude. That is, when working with T = 50, one must apply the tests mentioned above in order to obtain small errors. The AP and BP tests perform poorly at any specific break fraction. They are also sensitive to the break magnitude when working with sample sizes of 100, 250, and 500. The LS, Kapetanios (2005) and Perron (2006) tests are not superior to the BLS, ZA, and LP tests. In these tests, small break magnitudes do not perform well statistically. Their performances improve as the magnitude of the break increases.[2]

It can be observed from the test results of two break case[3] given in Figs. 2 and 3, the power of the LP test varies depending on the break magnitude. On the other hand, the LS test provides ambiguous results when the two breaks' magnitudes are near one another. It is clear from this evaluation that the LS test is sensitive to the break magnitude. LS predicts both break dates accurately when the second break magnitude is excessively large or extremely small compared to the first one. The tests conducted by Kapetanios (2005) and Perron (2006) provide good results in terms of how effectively they perform in determining the structural break date. For the two break cases, the test proposed by Bai and Perron (1998, 2003a, 2003b) has the poorest performance.[4]

In all figures, dark points on the surfaces of the separating hyperplanes (which can be represented as regression planes) correspond to the actual data points. MATLAB has plotted the 3D lines with reference to these points, enhancing the

[2] The comprehensive results of the performances of the related tests for one break in the mean case is given in Appendix 1.

[3] For the two break case, the analysis for T = 50 is not provided in the study because of the decreased sample size. The results can be provided upon request.

[4] The comprehensive results of the performances of the related tests for two break in mean case are exhibited in Appendix 2.

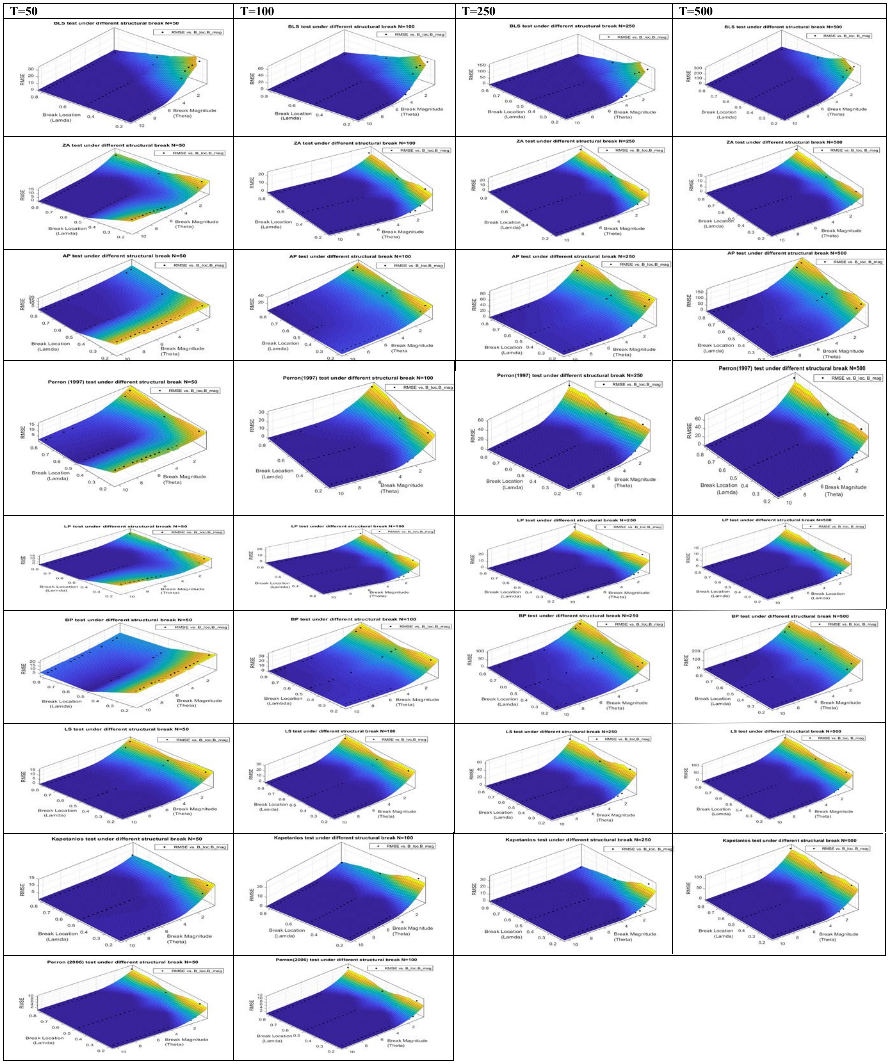

Fig. 1 RMSE results of a single break case for sample sizes of 50, 100, 250 and 500, respectively

visualization of the tables rather than the 3D data points themselves. For readers who are more interested in the technical details, the MATLAB 3D plot toolbox calculates the means of the actual points. Consequently, some data points lie on the separating hyperplanes, while others might not be visible as they are positioned beneath these separating hyperplanes. The blue and dark blue sections correspond to low root mean squared error (RMSE), whereas the light blue and yellow sections pertain to high RMSE. Also, in all Figures, N represents T.

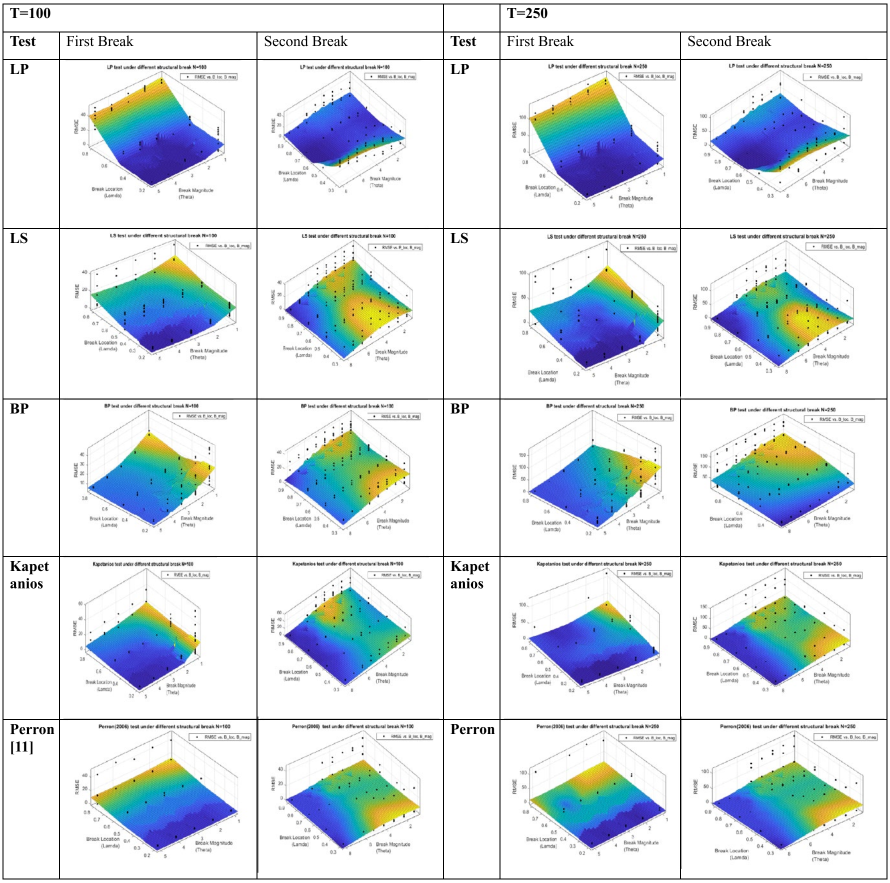

Fig. 2 RMSE results of two break case for sample sizes of 100 and 250

Summary table of the ranking of the tests employed for the simulation study is given in Table 2. The success of the test is ranked sequentially as 1, 2, 3, 4, and 1 indicates the best performed test.

As it is given in Table 2, the best performed test in detecting a single break is the ZA test. The least performed tests allowing a single break is both the AP and the Perron (1997) tests. Among the tests detecting two breaks, the Kapetanios (2005) and Perron (2006) tests exhibit the strongest performance. There is no significant difference in power between these two tests. The weakest performance is exhibited by the BP test among the structural break tests allowing two breaks.

In the empirical part of our study, all structural test applied to the series of real interest rate of the US and Australia. The performance of tests in the empirical part is determined based on the number of structural breaks that they can correctly capture. According to this criterion, the best performed unit root test

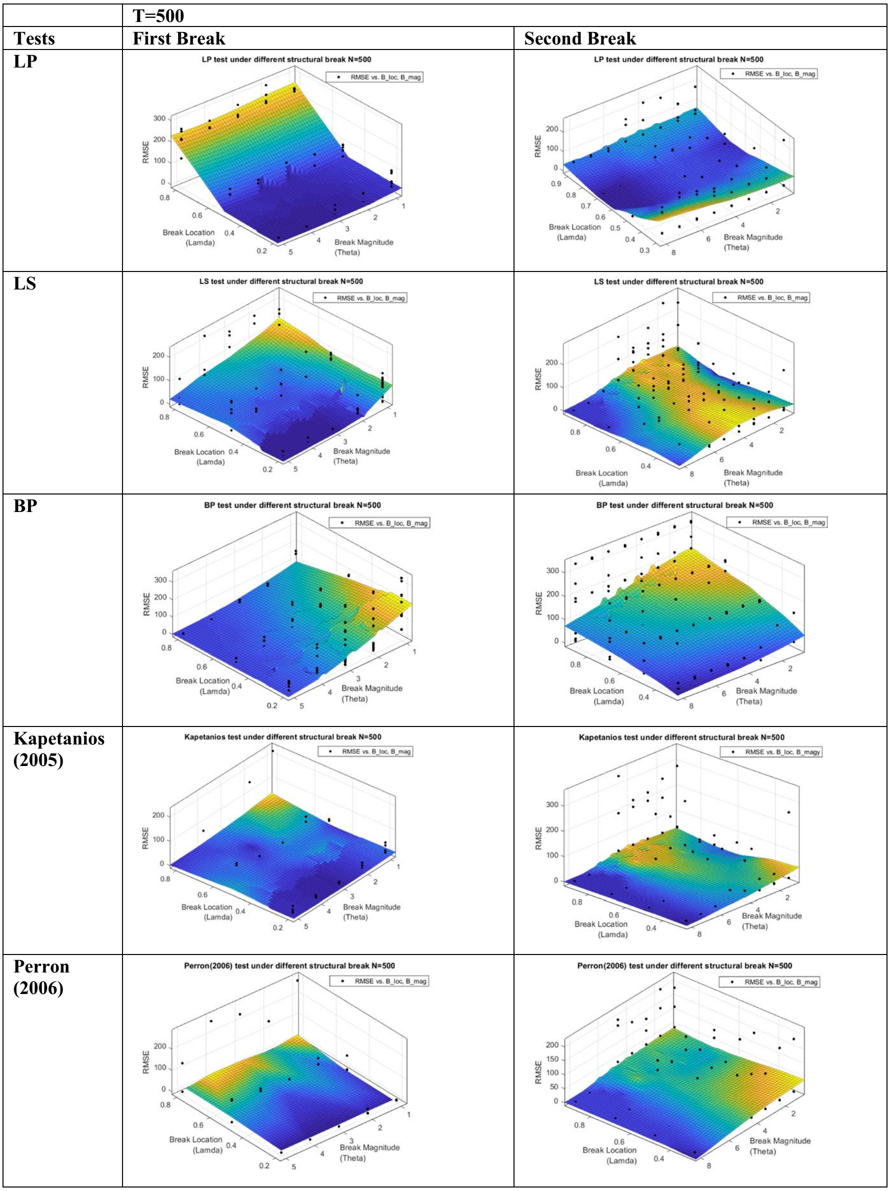

Fig. 3 RMSE results of two break case for sample size of 500

allowing multiple structural break test is found as the LP for the US, and the LP, LS, Kapetanios (2005) and Perron (2006) tests[5] for Australia.

[5] The detailed results of the empirical part are provided in the Sect. of 4.2.

Table 2 Summary table of simulation rankings of the tests

	Test	Simulation ranking	Empirical study ranking	
			US	Australia
Multiple structural break tests	LP	2	4	1
	BP	4	3	2
	LS	3	3	1
	Kapetanios (2005)	1	1	1
	Perron (2006)	1	2	1
Single structural break tests	ZA	1	–	–
	BLS	2	–	–
	AP	3	–	–
	Perron (1997)	3	–	–

5 Empirical Investigation and Newly Proposed Structural Break Significance Evaluation

This study used controlled data to compare different methods of detecting structural breaks through Monte Carlo simulation. The assumptions for the controlled data were as follows: (i) covariance stationarity, (ii) homoscedasticity or no conditional heteroskedasticity, (iii) absence of serial correlation, and (iv) a specified number of structural breaks. Under this framework, data containing a single structural break were generated and compared using various break detection methods. Subsequently, data incorporating multiple breaks were generated, and the analysis was repeated. However, it's important to note that these studies did not investigate how to find the optimal number of structural breaks in the data or what results would be obtained if the assumptions above were violated.

Additional factors beyond the assumptions listed above should be made when working with actual data (real world data). It is important, for example, to ascertain whether the deterministic part of the data exhibits a break in the mean or the trend. This problem has been discussed in detail in the economics literature for a variety of data sets and theories.

In Table 3, the type of structural break for each data set will be created, combined with insights from economic theory. For example, suppose only a structural break exists in the mean of the data. In that case, detecting a structural break in trend should not be pursued. If there is no information from economic theory, results from investigating trend breaks should be considered "spurious trend breaks". This is because economic theory does not provide an explanation for trend breaks, and statistically fitting a trend is only valid within the scope of data mining. Since the information provided by a structural break in mean and trend differs, considering both information is essential. Hence, determining and testing hypotheses based on insights from economic theory is crucial. Economic theory-based testing requirements for each data set are compiled in Table 3.

Table 3 Types of structural breaks in economic time series and hypothesis

Series/ Hypothesis	Mean	Trend	Mean + Trend
One price group			
Purchasing power parity (PPP)	The long-run purchasing power parity (PPP) hypothesis posits that exchange rates between two countries ought to adjust gradually over time in response to alterations in price levels, thereby maintaining parity in the purchasing power of a unit of currency across diverse nations	X	X
Quasi PPP	Temporary break in mean due to appreciation and depreciation across countries Taylor et al. discussed the concept of nonlinear mean-reversion in real exchange rates and its implications for solving the purchasing power parity puzzles. They highlighted faster adjustment speeds than previously recorded, indicating the presence of mean breaks due to real exchange rate shocks (Taylor et al., 2001)	In the context of quasi purchasing power parity, Hegwood and Papell (1998) discussed the implications of correctly estimating long-run equilibrium values for the real exchange rate. They emphasized that policy decisions and forecasts relying on PPP values depend on accurately estimating long-run equilibrium values, underscoring the importance of addressing structural breaks in trend analysis (Hegwood & Papell, 1998)	Mean + Trend
Real interest rate parity (RIP)	According to the calibrated neoclassical equilibrium growth model, a temporary increase in government expenditures persistently raises the real interest rate but eventually returns to its initial level (Baxter & King, 1993)	X	X
Uncovered interest parite (UIP)	Uncovered Interest Parity (UIP) theory posits that deviations from UIP will eventually be corrected, leading to exchange rates adjusting in a manner that aligns interest rate differentials with what is predicted by UIP	X	X
Real interest rate differantial (RID)	Research has shown that real interest rate differentials exhibit mean-reverting properties, indicating that they tend to revert to their equilibrium values over time (Baharumshah et al., 2005; Hoffmann & MacDonald, 2009; Smallwood & Norrbin, 2007). This behavior is characterized by long-memory dynamics, emphasizing the persistence but eventual return to a mean level (Baharumshah et al., 2005)	x	x

Table 3 (continued)

Series/ Hypothesis	Mean	Trend	Mean + Trend
Growth convergence			
Beta convergence	Beta convergence is defined as the negative relationship between the income growth and the initial income level. Beta-convergence refers to a process in which poor regions grow faster than rich ones and therefore catch on them. Jostova and Philipov (2005) propose a mean-reverting stochastic process for the market beta, indicating a connection between beta convergence and mean reversion.	x	x
Stochastic convergence	If countries have the same output predictions at a fixed time, there is stochastic convergence. In other words, if the difference in per capita incomes between two countries contains a unit root, there is no convergence. Sen and Singer (1993) discuss stochastic convergence in the context of economic growth and macroeconomics, providing insights into how stochastic processes can converge over time. Escobar et al. (2009) emphasize the mean-reverting nature of the covariance matrix, indicating that stochastic processes can indeed exhibit mean reversion in their dynamics	Linear trend	Linear trend

X denotes the absence of the related type of structural break. Numerous other economic testing hypotheses exist, but we have limited them as in the Table to not increase the table size

This study focuses on empirical research on real interest rate data; based on Table 3, we concentrate on the level break of real interest rate (RIP). In a study by Caporale and Grier (2005), real interest rate data for the United States were examined to investigate whether political changes cause structural breaks in the mean of the real interest rate series. The study considered two specific political-economic hypotheses: (1) political changes affect monetary policy, and (2) bureaucratic changes affect monetary policy. The structural break in the mean of the real interest rate resulting from these reasons was examined by applying the methodology proposed by Bai and Perron (1998). Following the study of Caporale and Grier (2005) and Mishra et al. (2023), we test the existence of a structural break in the mean in the Australian real interest rate series considering the same method.

Real interest rate data was collected from 1960:Q1–1999:Q4 for the US and 1970:Q1–2019:Q4. The nominal interest rate for the US was taken as a 3-month Treasury bill rate, while for Australia, it was used as a 90-day bank-accepted bill rate. Inflation rates were calculated as the percentage change in the Consumer Price Index (CPI) for both countries. Nominal interest rate data for Australia was obtained from the Main Economic Indicators (MEI)[6] database of the OECD. The 3-month treasury bill rate of the US and inflation rate for both the US and Australia were collected from the International Financial Statistics (IFS) database.

Real interest rate data is constructed by following Mishra et al. (2023) as $r_t = i_t - \pi_{t+1}$. r_t is the real interest rate at time t, i_t shows the nominal interest rate at the current year, and π_{t+1} is the inflation rate which is calculated as the percentage change in CPI over the time span between t and t + 1. r_t denotes the ex-post real interest rate.

This study excluded the Bai and Perron (1998) method from the Monte Carlo comparisons section for two main reasons. Firstly, the primary focus of this study is to compare break date detection methods tailored explicitly for unit root tests. Therefore, since the Bai and Perron (1998) method is not designed for unit root tests but rather for identifying structural breaks, it does not align with the study's objectives. Secondly, the Bai and Perron (1998) method is not designed for conducting unit root tests. Moreover, it suggests stationarity when performing structural break tests. Caporale and Grier's (2005) study discusses the assumption of stationarity of the Bai and Perron (1998) method. Still, they did not find it necessary to test it for US real interest rates. However, Mishra et al. (2023) addressed this issue by incorporating structural break detection into their simulation study, demonstrating that structural breaks were accounted for and the series was stationary.

In the section where the Bai-Perron process is explained in the Caporale and Grier (2005) method, the considerations to be taken into account and the assumptions under which the Bai and Perron (1998) method operates are summarized. These assumptions are categorized under three main headings: Stationarity, No Serial Correlation, and No Heteroskedasticity. Moreover, the Bai and Perron (1998)

[6] The 90-day bank accepted bill rate of Australia can be gathered from https://www.oecd.org/sdd/oecdmaineconomicindicatorsmei.htm, accessed on 3 February 2024.
The 3-month treasury bill rate of the US and inflation rate of the US and Australia data can be obtained from https://data.imf.org/?sk=4c514d48-b6ba-49ed-8ab9-52b0c1a0179b, accessed on 4 February 2024.

method is also used in two ways: sequential detection of the breakpoints and sample splitting. Caporale and Grier (2005) mentioned that both methods consistently found the same breakpoint dates, suggesting that the sequential break detection method might be evaluated as a more general structural breakpoint test. Therefore, a method that includes the safest repetition of each stage without intervention in every step needs to be proposed for this purpose.

Considering the restrictive assumptions in the Bai-Perron (1998) method in the Caporale and Grier (2005) and Mishra et al. (2023) studies, the Kapetanois (2005) methodology of sequential detection of the dominant breakpoint needs to be developed using the bootstrap method. Empirical research in this direction will provide a better opportunity to compare the results obtained in the Caporale and Grier (2005) and Mishra et al. (2023) studies.

5.1 A Methodology Proposed for Evaluating the Significance of Multiple Breaks in a Sequential Manner: A Bootstrap Approach

A new bootstrap approach will be proposed by following Namba (2017), ordinary/wild bootstrap (heteroskedasticity robust structural break detection), and combining it with the sieve bootstrap (nonstationarity under the null of the test) methods for assessing the significance of the F test for the nested hypothesis of sequential break detection. Bootstrap is a method used to estimate the distribution of a statistic or estimator by resampling data or a model estimated from data (Horowitz, 2001). As mentioned by Horowitz (2001), under appropriate conditions, bootstrap approaches approximate the distributions of statistics, the coverage probabilities of confidence intervals, and the rejection probabilities of hypothesis tests, which can be more accurate than first-order asymptotic distribution theories. Therefore, when it is difficult to calculate the asymptotic distribution of a statistic or estimator, critical values obtained using the bootstrap method will be more accurate and represent a good approximation of the asymptotic distribution in small samples. This topic is also stated by Horowitz (2001), noting that "Bootstrap generally provides more accurate results than first-order asymptotic approaches, but does not require the algebraic complexity of higher-order expansions".

Caporale and Grier (2005) utilized the Bai and Perron (1998) approach in their methodology, with its crucial assumption being stationarity. In this context, Mishra et al. (2023) believed they addressed the issue Caporale and Grier (2005) overlooked by conducting all structural break tests containing unit root tests in our simulation section. However, the proposed methods did not apply consecutive and sample-splitting unit root tests. Caporale and Grier (2005) also mitigated the assumed absence of serial correlation and changing variance in the BP method by using Andrews's (1991) robust standard errors. Therefore, considering they thought they addressed so many issues without testing or controlling, they actually inferred results with inherent weaknesses. Hence, a more straightforward and more robust method should be proposed.

Namba (2017) ordinary/wild bootstrap methods mitigated the size distortion arising from changing variance in the Chow test methodology. Likewise, the sieve bootstrap methodology was used to test the unit root null hypothesis (Uçar and Omay, 2009). These two bootstrap approaches will generate critical values for robust structural break testing against dynamics such as unit roots, changing variance, and serial correlation under the null hypothesis, where the series has no structural break. Therefore, robust critical values against the mentioned conditions will be produced when an alternative structural break exists. Thus, multiple structural break tests will be performed more simply and robustly compared to the Caporale and Grier (2005) methodology. The proposed bootstrap methodology is outlined below:

Estimate the M break Model

$$y_t = \mu_1 + \mu_2 d_2 + e_t,\ \left(d_2 = 0 \text{ if } \tau_1 < \text{t}, d_2 = 1 \quad \text{if } \tau_1 > \text{t}\right)$$

$$y_t = \mu_1 + \mu_2 d_2 + \mu_3 d_3 + e_t\left(d_2 = 1 \text{ if } \tau_1 < \text{t} \leq \tau_2,\ d_3 = 1\text{if } \tau_2 > \text{t}\right)$$

$$y_t = \mu_1 + \mu_2 d_2 + \mu_3 d_3 + \mu_4 d_4 + e_t\left(d_2 = 1 \text{ if } \tau_1 < \text{t} \leq \tau_2\ d_3 = 1 \text{ if } \tau_2 < \text{t} \leq \tau_3\ d_3 = 1 \text{ if } \tau_3 > \text{t}\right)$$

$$y_t = \mu_1 + \sum_{i=2}^{M} \mu_i d_i + e_t\left(d_2 = 1 \text{ if } \tau_1 < \text{t} \leq \tau_2,\ d_3 = 1 \text{ if } \tau_2 < \text{t} \leq \tau_3, \cdots, d_M = 1 \text{ if } \tau_M > \text{t}\right)$$

M is the number of break.

$$F = \frac{SSR_r - SSR_{ur}/df}{SSR_{ur}/df}$$

(i) The following OLS regression is considered, and we imposed the null of no break.

$$y_t = \hat{\beta}_1 + \hat{\beta}_2 y_{t-1} + \hat{\varepsilon}_t^0$$

(ii) The null of no unit root is imposed to generate samples of residuals (Basawa et al., 1991). Error terms are estimated as below:

$$\Delta\hat{\varepsilon}_t^1 = \Delta y_t - \tilde{\hat{\beta}}_1 - \tilde{\hat{\beta}}_2 \Delta y_{t-1}$$

(iii) Stine (1987) offers that the residuals must be centered with

$$\Delta\hat{\varepsilon}_t^1 = \Delta\hat{\varepsilon}_t^1 - \frac{1}{(T-1-2)} \sum_{t=1+2}^{T} \Delta\hat{\varepsilon}_t^1$$

(iv) We first generate stationary bootstrap residuals recursively from

$$\varepsilon_t^* = \sum_{t=1}^{T} \Delta\varepsilon_t^1$$

(v) Following Namba (2017) the below data generation is obtained:

$$y_{i,t}^* = \hat{\beta}_1 + \hat{\beta}_2 y_{t-1} + \varepsilon_t^*$$

(vi) The bootstrap statistics are computed for each bootstrap replication by running the regressions.

$$y_t^* = \mu_1 + \mu_2 d_2 + u_t, \left(d_2 = 0 \text{ if } \tau_1 < \text{t}, d_2 = 1 \text{ if } \tau_1 > \text{t}\right)$$

$$y_t^* = \mu_1 + \mu_2 d_2 + \mu_3 d_3 + u_t\left(d_2 = 1 \text{ if } \tau_1 < \text{t} \leq \tau_2,\ d_3 = 1 \text{ if } \tau_2 > \text{t}\right)$$

$$y_t^* = \mu_1 + \mu_2 d_2 + \mu_3 d_3 + \mu_4 d_4 + u_t\left(d_2 = 1 \text{ if } \tau_1 < \text{t} \leq \tau_2\ d_3 = 1 \text{ if } \tau_2 < \text{t} \leq \tau_3\ d_3 = 1 \text{ if } \tau_3 > \text{t}\right)$$

$$y_t^* = \mu_1 + \sum_{i=2}^{M} \mu_i d_i + u_t\left(d_2 = 1 \text{ if } \tau_1 < \text{t} \leq \tau_2,\ d_3 = 1 \text{ if } \tau_2 < \text{t} \leq \tau_3, \cdots, d_M = 1 \text{ if } \tau_M > \text{t}\right)$$

*M*th break.

(vii) Obtain the empirical F_i^* test and use for construction of bootstrap or empirical density function:

$$F_i^* = \frac{SSR_r - SSR_{ur}/df}{SSR_{ur}/df} \sim EDF\left(MomentsF_i^*, i = 1, ..., B\right)$$

Empirical distribution of these statistics is produced by 2000 replications, thus, their p-values are generated.

A consecutive break date is obtained sequentially for each break, from the dominant break to the recessive break. Therefore, this sequential process incurs a nested F test hypothesis testing starting from 1 break to no break null end with M-1 break to M break alternative hypothesis.

Remark: We designed this 7-stage process under the most general assumptions; heteroscedasticity and nonstationarity. With each break added in each stage, the empirical distribution function changes, and after a stage, the remaining series should exhibit stationary properties. Hence, the empirical distribution function must incur these features. Therefore, continuing with the bootstrap algorithm with $\hat{\varepsilon}_t^0$ is assumed to be a more accurate approach. Thus, skipping steps 2, 3, and 4 and obtaining $\hat{\varepsilon}_t^*$ in step 5 with $\hat{\varepsilon}_t^0$ instead of $\hat{\varepsilon}_t^1$ is better to generate the suitable empirical distribution function.

5.2 Comparison of the Newly Proposed Bootstrap Approach and Caporale and Grier (2005) Method

Depending on this proposed bootstrap algorithm, we have obtained the following results for the US between 1960:Q1–1999:Q4, and for Australia 1970: Q1–2019: Q4.

Caporale and Grier's (2005) approach identifies four breaks, whereas our proposed bootstrap method reveals three statistically significant break dates. Although the first three break dates coincide with those identified by Caporale and Grier (2005), the fourth statistically significant break in their method is recognized as the fifth dominant break in our proposed method. However, this fifth dominant break is statistically insignificant.

The method proposed by Caporale and Grier (2005) differs from the sequential approach of the Bai and Perron (1998) method. Their method divides the sample into different blocks, and the process is repeated. However, this process has inconsistencies within itself. Primarily, the Bai and Perron (1998) method is applied to stationary processes. Therefore, the initial lack of stationarity in the series affects the distribution evaluated for the initial result. Caporale and Grier (2005) described the process they used as follows: *"A sequential procedure can also be used to select the number of breaks in which, if an initial break is found [based on the initial SupFt(l) test], the sample is then divided into subgroups at the break point, and the same parameter constancy test is then performed on the subsamples. The partitioning of the subsamples continues until the parameter constancy test fails to reject the null"*. Hence, since the stationary test is not performed for each sample separated, it is doubtful to obtain the correct results. Additionally, the assumption set used in the Bai and Perron (1998) method states, "*Serial correlation in the errors and nonconstant error variances within and between segments*". Caporale and Grier (2005) explained that they corrected the standard errors for each sub-group with Andrews's (1991) and Newey and West's (1987) methods. We designed the bootstrap method to cover all described scenarios, ensuring its robustness against these cases by incorporating *F*-test critical values derived from the empirical distribution function. Furthermore, Caporale and Grier (2005) state that *"determining the number and location of the breaks, 90, 95, and 99% confidence intervals for the break dates are provided. The derivations of these values are explained in Bai and Perron (2000, pp. 11–13) and rest on using a novel asymptotic framework in which the magnitudes of the shifts converge to zero as the sample size*". We do not consider the confidence intervals obtained here as reliable. For instance, for 1967:Q2, a 95% confidence interval is given as 1966:Q1 and 1969:Q2. In this confidence interval, four lower and seven upper quartiles are drawn.

Table 4 US break dates and the Bootstrap F -test critical values

Cap&Gr	1980:Q3	1973:Q1	1986:Q2		1967:Q2	
BP	1980:Q1	1971:Q4	1985:Q1		1966:Q1	
Confidence intervals	1980:Q3	1973:Q1	1986:Q2		1967:Q2	
	1981:Q1	1973:Q2	1987:Q3		1969:Q2	
	(1,1)	(4,0)	(4,4)		(4,7)	
Newly proposed bootstrap method						
Break date	1980:Q3	1973:Q2	1986:Q3	1981:Q2	1967:Q2	1994:Q4
F test	27.681 (0.000)	20.059 (0.000)	17.798 (0.000)	1.592 (0.899)	1.242 (0.609)	0.923 (0.758)
%1	23.517	13.652	6.749	6.914	3.163	2.608
%5	21.859	11.617	5.560	5.715	2.420	2.146
%10	18.384	10.403	5.057	5.068	2.194	1.935

Cap&Gr indicates the Caporale and Grier (2005) estimated break dates obtained by their proposed method

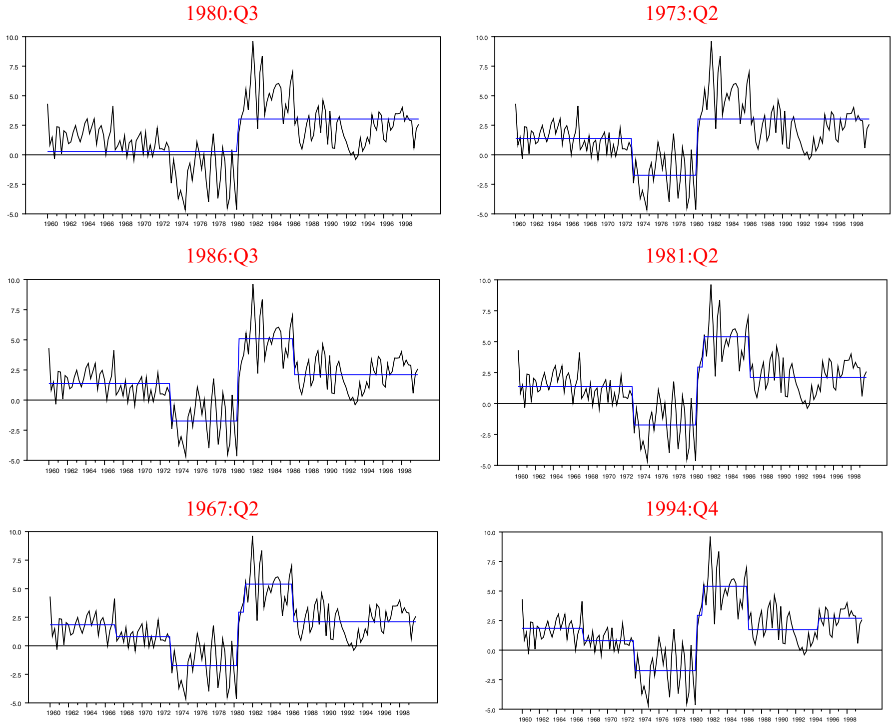

Fig. 4 Sequential break date estimation and the locations of break for the US

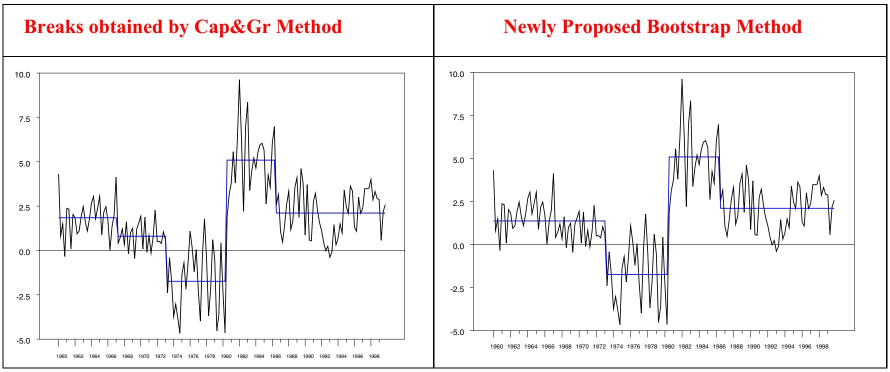

Fig. 5 Comparison of Cap&Gr method and newly proposed bootstrap method

For the other three breaks, 1971:Q4–1973:Q1–1973:Q2 (4,0), 1980:Q1–1980:Q3–1981:Q1 (1,1), 1985:Q1–1986:Q2–1987:Q3 (4,4) are determined. As seen from Table 4, the dominant break is 1980:3 with 1% significance level. Similarly, the Bai and Perron (1998) method also provides a very narrow confidence interval. Providing a narrow confidence band indicates the

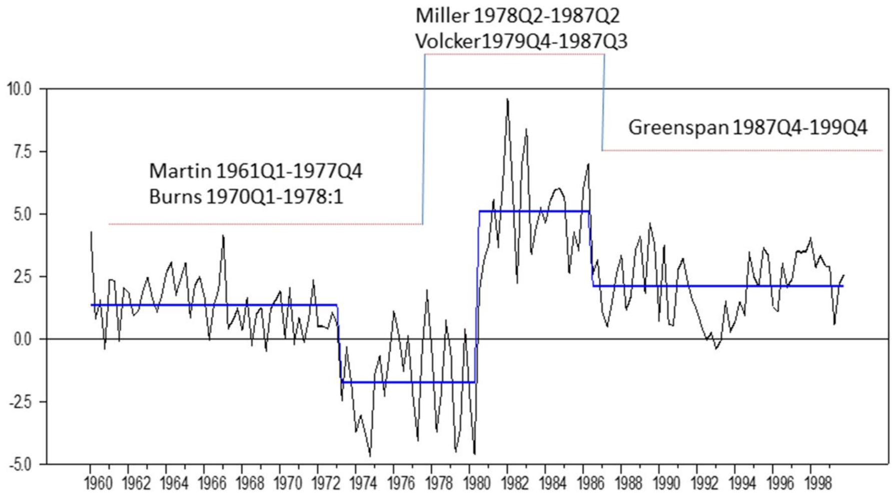

Fig. 6 Important break dates depending on the bureaucracy changes in the US

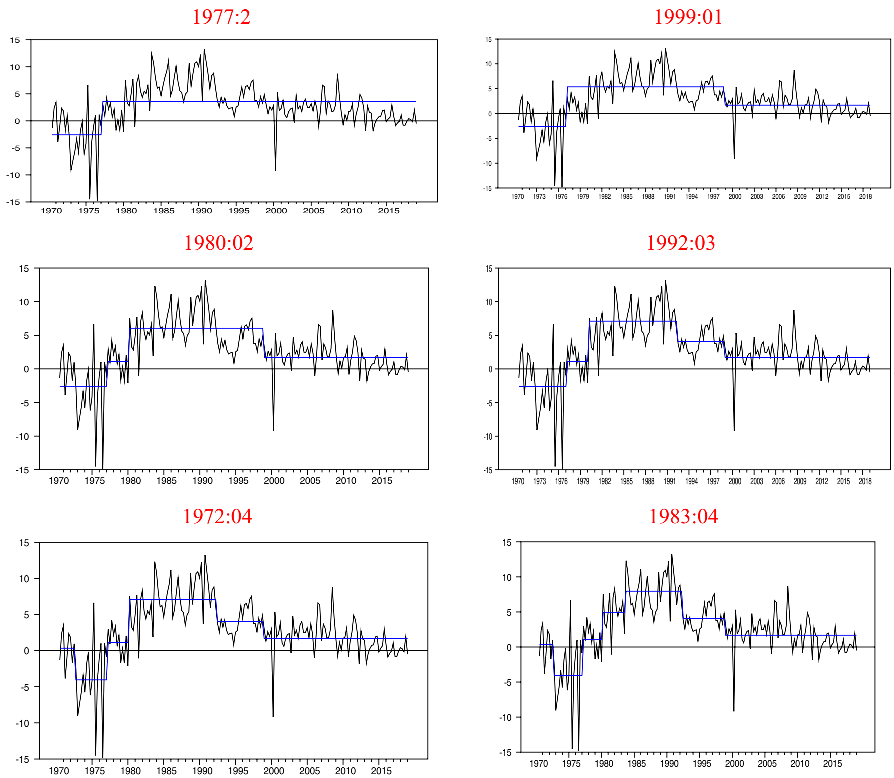

Fig. 7 Sequential break date estimation and the locations of break for the US

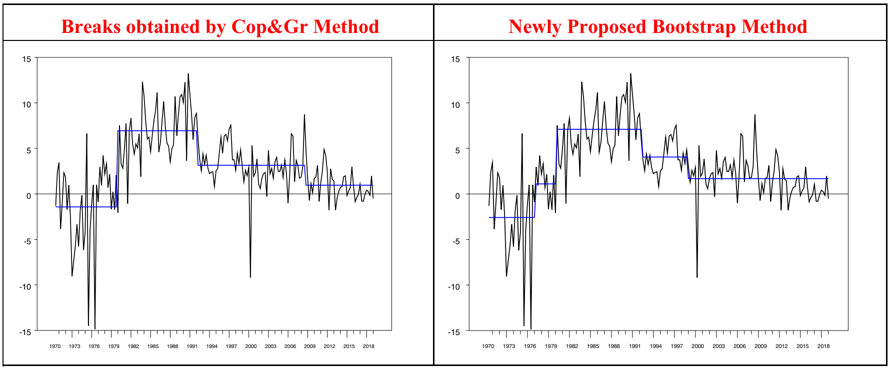

Fig. 8 Comparison of Cap&Gr method and newly proposed bootstrap method

high significance of the structural break here. While the other two breaks remain relatively narrow, 1966:Q1–1967:Q2–1969:Q2 (4,7) is statistically significant in a wide band interval, indicating the dubiousness of the result due to the methodological shortcomings explained earlier. Since the Bai and Perron (1998) method is conducted without considering many critical assumptions, there is a high probability that the critical values or confidence intervals generated may be biased. Our proposed bootstrap method rejected the 1967:Q2 break found in the fifth position (0.609) at a 61% statistical significance level. Therefore, it was concluded that there is no significant break, and the null hypothesis of no break cannot be rejected. At this point, it is worth mentioning that the Kapetanios (2005) test also applies this method as a break selection criterion and avoids arbitrary break selection. Furthermore, the bootstrap method proposed here can access the more complex form of the asymptotic distribution that Kapetanios (2005) could not reach. In this sense, developing a similar bootstrap algorithm can also suitable for Kapetanios (2005) unit root test (The unique feature of the newly proposed test can be accurately followed through following the Figs. 4, 5, 6, 7, 8).

To provide additional insights into the proposed bootstrap method based on the *F*- test table, we have added the significance levels of 10%, 5%, and 1% obtained from the *F*- test alongside the bootstrap statistics in small font. The bootstrap critical values obtained and the *F*-test critical values converge closely after each sequentially obtained break. Explaining this phenomenon would be useful in demonstrating the consistency of the proposed bootstrap method. The proposed bootstrap method is robust against nonstationarity, heteroscedasticity, and serial correlation. For instance, upon finding the first break, if we had referred to the standard *F*-test table, we would have concluded that the break date is significant at the 1% level, with F(1,195,%10) = 6.81, whereas the computed test value is 33.231. Following the structural break found in March 1992, there is a divergence in the significance levels, with the *F*-test indicating significance at the 1% level for the existence of a break. In contrast, our proposed method suggests significance at the 10% level.

Regarding the break date in April 1972, our proposed method does not find it statistically significant at the traditional significance level, while the *F*-test would

Table 5 Australia break dates and the Bootstrap *F*- test critical values

Cap&Gr			1980:Q1	1992:Q2		
BP confidence intervals			1979:Q3 1980:Q1 1981:Q1 (1,3)	1990:Q4 1992:Q2 1994:Q4 (5,9)		
Break date	1977:Q2	1999:Q1	1980:Q2	1992:Q3	1972:Q4	1983:Q4
F test	33.231 (0.000)	18.156 (0.000)	6.887 (0.010)	3.755 (0.071)	2.457 (0.191)	1.746 (0.274)
%1	22.311 (6.81)	11.960 (4.71)	6.676(3.88)	6.340 (3.41)	5.995 (3.11)	4.981 (2.90)
%5	18.384	9.344	5.073	4.705	3.963 (2.26)	3.352 (2.14)
%10	16.360	8.229	3.852	3.661	3.119	2.784 (1.73)
Cap&Gr			2008:Q4			
BP Confidence intervals			2006:Q4–2008:Q4–2013:Q3 (7,19)			
Break date	2012:Q2	1976:Q4				
F test	1.177 (0.393)	1.177 (0.330)				
%1	3.960 (2.73)	4.322 (2.60)				
%5	3.119 (2.05)	2.900 (1.98)				
%10	2.121 (1.72)	2.347 (1.67)				

Cap&Gr indicates the Caporale and Grier (2005) estimated break dates obtained by their proposed method

Table 6 Comparison of all structural break tests

	LP	LS	BP_5	Kap_5	Perron_5	Cap&Gr
US	1980:Q3*	1973:Q4*	1973:Q3*	1980:Q3*	1973:Q2	1967:Q2
1960:Q1–1999:Q4	1994:Q1	1980:Q3*	1975:Q1	1973:Q2*	1980:Q2*	1973:Q1*
			1979:Q4	1986:Q3*	1986:Q1*	1980:Q3*
			1981:Q1*	1981:Q2	1990:Q1	1986:Q2*
			1983:Q2	1967:Q2	1993:Q4	
Success rate w.r.t. original model	1/4	2/4	2/4	4/4	3/4	4/4
Success rate w.r.t. Bootstrap model	1/3	2/3	2/3	3/3	2/3	3/3
Australia	1980:Q1**	1980:Q1**	1972:Q3	1977:Q2*	1979:Q4	1980:Q1**
1970:Q3–2019:Q1	1992:Q2***	1992:Q2***	1974:Q1	1999:Q1*	1981:Q2**	1992:Q2***
			1977:Q1	1980:Q2**	1991:Q4	2008:Q4
			1999:Q2	1992:Q3***	2000:Q1	
Success rate w.r.t. original model	2/3	2/3	0/3	2/3	2/3	3/3
Success rate w.r.t. Bootstrap model	2/4	2/4	0/4	4/4	1/4	2/4

*,**, and *** denote %1, %5 and %10 significance level of break date estimates depending on the bootstrap Kapetanios type F-test. We accepted errors up to two quarters. w.r.t. indicates "with respect to"

have considered it significant at the 5% level. Hence, an additional break date would have been unnecessarily added. As nonstationarity, heteroscedasticity, and serial correlation decrease in each sequence, the empirical distribution function and traditional F-test asymptotic values converge. Therefore, the results provided for Australia Table 5 demonstrate the consistency of the proposed bootstrap method.

The proposed BP method for Australia has found narrow confidence intervals for the breaks at 1979:Q3–1980:Q1–1981:Q1 (1,3), 1990:Q4–1992:Q2–1994:Q4 (5,9), and 2006:Q4–2008:Q4–2013:Q3 (7,19). Using our proposed method, we identified statistically significant breaks at 1977:Q2, 1999:Q1, 1980:Q2, and 1992:Q3. Out of the four breaks we identified, only two align with Mishra et al. (2023)'s breaks, while the break given for 2008:Q4 as having low significance (wide confidence band) was not obtained even after increasing the number of breaks to eight. As evident from both studies, our proposed bootstrap method has performed excellently in determining the optimal number of breaks within the context of these two examples.

5.3 Empirical Research: Remaining All Tests Which are Compared in Sect. 3

In the empirical research part of our study, the results obtained are compared with the findings from the Monte-Carlo study section. We confirmed that there would be a break in the mean from two sources: political party changes or changes in the

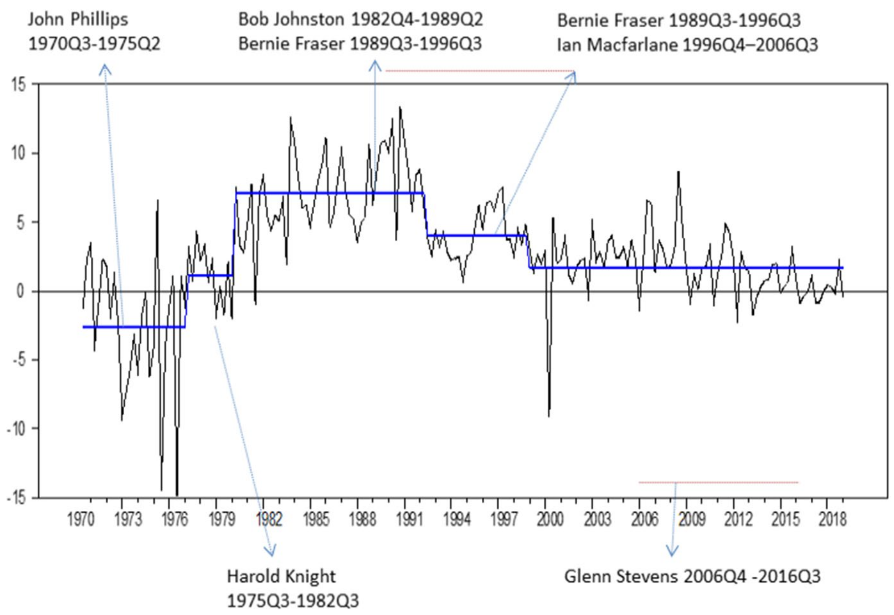

Fig. 9 Important break dates depending on the bureaucracy changes in Australia

central bank governor, as corroborated by the study of Caporale and Grier (2005). We have also introduced a bootstrap methodology to further confirm these break dates.

Table 6 shows that Kapetanios (2005) has the highest success rate in capturing possible break dates. When looking at Figs. 6 and 9, it becomes apparent that rather than political party changes, Federal Reserve chair changes, i.e., bureaucratic shifts, have played a significant role in determining the real interest rate mean values. The mean value change of the real interest rate is nearly one-to-one, corresponding to each change in the Federal Reserve chair.

While the sequential implementation of the Bai and Perron (1998) method yields similar results to the partitioned sampling approach in the US data set, it shows zero performance in the Australian dataset, performing worse than all other methods. The LP and LS methods are two methods that predict up to two structural breaks. The LP correctly predicts one of the two breaks for the US and both for Australia. The LS correctly predicts both breaks for both countries. However, due to limitations in these two methods, it would be more appropriate to compare the the Bai and Perron (), Kapetanios (2005), and Perron (2006).

The BP achieves two out of four successes with the sequential method for the US, while for Australia, it achieves zero successes. Kapetanios (2005) captures four out of four successes for the US and two out of three successes for Australia. Perron (2006) achieves three out of four successes for the US and two out of three for Australia. Although the break date estimations in the original articles may be questioned, our proposed bootstrap method guarantees optimal significant break date selection and provides a benchmark for comparison.

The empirical study results are consistent with the results obtained in Monte Carlo simulations. The general conclusion can be derived as Kapetanios' (2005) sequential break estimation performed better than all other tests included in the study. Perron's (2006) multiple structural break method performed weaker than Caporale and Grier's (2005) suggested sample splitting BP test. The sequential BP method by Bai and Perron (1998) performs even worse than the LP and LS methods in finding two breaks.

6 Concluding Remarks

In this study, the main aim is to compare the performances of the structural breaks and their break dates. Zivot and Andrews (1992), Baneerje, Lumsdaine, and Stock (Banerjee, 1992), Andrews and Ploberger (1994), Perron (1997) tests allowing a single structural break, and Lumsdaine and Papell (1997), Bai and Perron (Bai & Perron, 2003a), Lee and Strazicich (2003), Kapetanios (Kapetanios, 2005) and Perron (Perron, 2006) tests allowing multiple structural breaks are applied to the randomly generated series. In order to determine the performances of the tests and compare them, the Monte Carlo simulation experiments with 1000 simulation trials are applied to the randomly generated series with a normal distribution for the one break in mean and two break in mean cases. For the one break case, the break fractions have been placed at the beginning, in the middle and at the end of the series. The performances of the tests depending on the location were obtained. In the two break cases, the break fractions were calculated in the same way. For the second break, the break fraction was gradually removed from the first one. It is found out that the sensitivity of the tests are increased as break magnitude increases. Considering both cases, a common conclusion from the simulations is that the Kapetanios (2005) and Perron (2006) tests performed better than the other tests.

To conduct the empirical part of the study more robustly, it was necessary to determine the optimal statistically significant number of breaks. The main problem in this regard, within the framework of the Bai and Perron (1998) method, assumes stationarity for the structural break test. Additionally, it is known that periods of high volatility, i.e., increased heteroscedasticity or unequal variance in sample segments where breaks occur, would significantly affect the reliability of F-test statistics. We proposed a new bootstrap methodology to address these two fundamental issues, thereby developing a robust bootstrap algorithm that accounts for heteroscedasticity and nonstationarity under the null. Through the empirical study, we determined that the proposed bootstrap algorithm generates data in a structure that accurately represents an empirical distribution function, thereby assisting in obtaining accurate critical values of the challenging asymptotic distribution.

Furthermore, it is evident that the sequential test stages converge towards a traditional F distribution under the null by effectively eliminating nonstationarity, heteroscedasticity, and serial correlation. Modeling each break leaves behind a smoother series closer to white noise, ensuring stationarity, homoscedasticity, and no serial correlation structure after each break estimation. This is further supported by the fact that the generated bootstrap crucial values converge to the F test values obtained under these assumptions. This phenomenon is observed in Tables 4 and 5.

Moreover, we can use the bootstrap *F*-test for the pre-test of Kapetanios (2005) unit root test. Hence, we can use statistically significant breaks for the unit root testing. It would be more meaningful if the optimal number of statistically significant breaks leads to a stationarity conclusion. Thus, we believe that we have made a significant contribution to the literature by proposing a bootstrap algorithm that allows for more reliable handling of pre-testing with sequential break detection methods.

The study's main objective was to measure the performance of structural break tests conducted within unit root tests by generating controlled structural breaks under the conditions of stationarity, homoscedasticity, and no serial correlation. However, evaluation problems encountered in the empirical section and performance criterion have led to new proposals and addressing existing deficiencies in the literature. Since the proposed methods were developed based on the empirical study, there is considerable benefit in confirming them through broader Monte Carlo simulations and conducting theoretical investigations. The results obtained within the scope of the empirical research are consistent and align with expectations. Therefore, it can be concluded that the article's primary objective, comparing break detection methods within unit root tests, has been achieved. It has been concluded that the sequential Chow test-type break detection methods proposed by Kapetanios (2005) yield better results.

Appendix 1: Comparison of tests in the case of a single break for T = 50, 100, 250, 500

T = 50				T = 100			
Break magnitude	Break location			Break magnitude	Break location		
	0.2	0.5	0.8		0.2	0.5	0.8
0.5	Perron (2006)	Perron (2006)	BLS	0.5	Perron (2006)	Perron (2006)	BLS
1	Perron (2006)	Perron (2006)	Kapetanios	1	Perron (2006)	Perron (2006)	Kapetanios
1.5	Perron (2006)	Perron (2006)	Kapetanios	1.5	ZA. LP	ZA. LP	ZA. LP
2	Perron (2006)	ZA. LP	ZA. LP	2	ZA. LP	ZA. LP	ZA. LP
2.5	Perron (2006)	ZA. LP	ZA. LP	2.5	ZA. LP	ZA. LP	ZA. LP
3	Perron (2006)	ZA. LP	ZA. LP	3	ZA. LP	ZA. LP	ZA. LP
3.5	Perron (2006)	ZA. LP	ZA. LP	3.5	ZA. LP	ZA. LP	ZA. LP
4	LS	ZA. LP	ZA. LP	4	ZA. LP	ZA. LP	ZA. LP
4.5	LS	ZA. LP	ZA. LP	4.5	ZA. LP. LS	ZA. LP	ZA. LP
5	LS	ZA. LP	ZA. LP	5	ZA. LP. LS	ZA. LP	ZA. LP. LS
5.5	LS	ZA. LP	ZA. LP. LS	5.5	ZA. LP. LS	ZA. LP	ZA. LP. LS
6	LS	ZA. LP	ZA. LP. LS	6	ZA. LP. LS	ZA. LP	ZA. LP. LS
6.5	LS	ZA. LP	ZA. LP. LS	6.5	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS
7	LS	ZA. LP	ZA. LP. LS	7	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS
7.5	LS	ZA. LP	ZA. LP. LS	7.5	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS
8	LS	ZA. LP	ZA. LP. LS	8	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS
8.5	LS	ZA. LP	ZA. LP. LS	8.5	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS
9	LS	ZA. LP. LS	ZA. LP. LS	9	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS
9.5	LS	ZA. LP. LS	ZA. LP. LS	9.5	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS
10	LS	ZA. LP. LS	ZA. LP. LS	10	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS

T = 250				T = 500			
Break magnitude	Break location			Break magnitude	Break location		
	0.2	0.5	0.8		0.2	0.5	0.8
0.5	Perron (2006)	Perron (2006)	Kapetanios	0.5	Perron (2006)	Perron (2006)	Perron (2006)
1	ZA. LP	ZA. LP	Kapetanios	1	ZA. LP	ZA. LP	ZA. LP
1.5	ZA. LP	ZA. LP	ZA. LP	1.5	ZA. LP	ZA. LP	ZA. LP
2	ZA. LP	ZA. LP	ZA. LP	2	ZA. LP	ZA. LP	ZA. LP
2.5	ZA. LP	ZA. LP	ZA. LP	2.5	ZA. LP	ZA. LP	ZA. LP
3	ZA. LP	ZA. LP	ZA. LP	3	ZA. LP	ZA. LP	ZA. LP
3.5	ZA. LP	ZA. LP	ZA. LP	3.5	ZA. LP	ZA. LP	ZA. LP
4	ZA. LP	ZA. LP	ZA. LP	4	ZA. LP. LS	ZA. LP	ZA. LP
4.5	ZA. LP. LS	ZA. LP	ZA. LP. LS	4.5	ZA. LP. LS	ZA. LP	ZA. LP. LS
5	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS	5	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS
5.5	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS	5.5	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS
6	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS	6	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS
6.5	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS	6.5	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS
7	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS	7	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS
7.5	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS	7.5	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS
8	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS	8	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS
8.5	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS	8.5	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS
9	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS	9	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS
9.5	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS	9.5	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS
10	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS	10	ZA. LP. LS	ZA. LP. LS	ZA. LP. LS

Appendix 2: Comparison of tests for two breaks case for T = 100, 250, 500

T = 100															
Break location															
Break magnitude		0.2								0.5				0.8	
		0.3		0.5		0.7		0.9		0.7		0.9		0.9	
1	1	LP	LP	LP	LP	LP	LP	LP	BP	BP	BP	LP	LP	BP	LP
	2	P	P	LP	LP	LP	LP	K	P	LP	LP	P	K	P	P
	3	P	P	LP	K	LP	LP	K	K	LP	LP	P	K	P	P
	4	P	K	LP	LP	LP	LP	K	K	LP	LP	LS	K	P	LS
	5	P	P	LP	LP	LP	LP	P	K	LP	LP	P	LS	P	LS
	6	P	P	LP	LP	LP	LP	P	LS	LP	LP	P	LS	P	LS
	7	P	P	LP	LP	LP	LP	P	LS	LP	LP	P	LS	P	LS
	8	P	P	LP	LP	LP	LP	P	LS	LP	LP	P	LS	P	LS
2	1	LP	P	LP	LP	LP	LP	LP	LP	LP	LP	LP	P	P	P
	2	LP	LP	LP	LP	LP	LP	LP	P	LP	LP	LP	BP	LP	K
	3	LP	P	LP	P	LP	LP	LP	K	LP	LP	LP	P	P	P
	4	LP	P	LP	LP	LP	LP	LP	K	LP	LP	LP	K	P	P
	5	P	K	LP	LP	LP	LP	LP	K	LP	LP	LP	K	P	LS
	6	P	K	LP	LP	LP	LP	LP	K	LP	LP	LP	K	P	LS
	7	P	K	LP	LP	LP	LP	LP	K	LP	LP	LP	LS	P	LS
	8	P	K	LP	LP	LP	LP. LS	LP	K	LP	LP	LP	LS	P	LS

T = 100															
Break location															
Break magnitude		0.2								0.5				0.8	
		0.3		0.5		0.7		0.9		0.7		0.9		0.9	
3	1	LP	P	LP	LP	LP	LP	LP	K	LP	LP	LP	K	K	P
	2	LP	P	LP	LP	LP	LP	LP	K	LP	LP	LP	LP	K	P
	3	LP	K	LP	BP	LP	LP	LP	P	BP	BP	LP	LP	BP	K
	4	LP	P	LP	K	LP	K	LP	K	LP	LP	LP	LP	K	P
	5	LP	P	LP	LP	LP	LP	LP	K	LP	LP	LP	LP	K	LS
	6	LP	K	LP	LP	LP	LP	LP	K	LP	LP	LP	LS	P	K
	7	LS	LS	LP	LP	LP	LP	LP	K	LP	LP	LP	LS	K	LS
	8	P	P	LP	LP	LP	LP	LP	K	LP	LP	LP	LS	P	LS. K
4	1	LP	LS	LP. LS	LP	LP. LS	LP	LP	K	LP	LP	LP	K	K	K
	2	LP	P	LP	LP	LP. LS	LP	LP	K	LP	LP	LP	K	K	P
	3	LP	P	LP	LP	LP	P	LP	P	LP	LP	K	P	P	P
	4	LP	K	LP	LP	LP	P	LP	P	BP	BP	BP	LP	BP	BP
	5	LP. LS	P	LP	LP	LP	LP	LP	P	LP	LP	K	P	BP	LS
	6	LP. LS	P	LP	LP	LP. LS	LP	LP	K	LP	LP	LP	K	K	P
	7	LP. LS	P	LP	LP	LP	LP	LP	K	LP	LP	LP	K	K	P
	8	LP. LS	LS	LP. LS	LP	LP. LS	LP	LP	K	LP	LP	LP	LS	P	LS

T = 100															
Break location															
Break magnitude		0.2								0.5				0.8	
		0.3		0.5		0.7		0.9		0.7		0.9		0.9	
5	1	LP. LS	LS	LP. LS. K	LP. LS	LP. LS	LP	LP	LS	LP	LP	LP. LS	LS	K	K
	2	LP. LS. K	LS	LP. LS. K	LP	LP. LS	LP	LP	K	LP	LP	LP. LS	K	K	P
	3	LP. LS. K	P	LP. LS	LP	LP	LP	LP	K	LP	LP	LP	P	K	P
	4	LP	P	LP	LP	LP	P	LP	P	LP	P	LP	P	K	P
	5	LP	LP	LP	BP	LP	LP	LP	P	LP	BP	BP	LP. P	BP	BP
	6	LP	P	LP	LP	LP	LP	LP	P	LP	P	LP	P	BP	P
	7	LP. LS	P	LP. LS	LP	LP. LS	LP	LP	K	LP	LP	LP	K	K	P
	8	LP. LS	P	LP. LS	LP	LP. LS	LP	LP	K	LP	LP	LP	K	K	P. K

T = 250															
Break location															
Break magnitude		0.2								0.5				0.8	
		0.3		0.5		0.7		0.9		0.7		0.9		0.9	
1	1	P	K	LP	LS	LP	P	LP	LS	LP	LP	LP	BP	BP	LP. K
	2	P	P	LP	LP. K	LP	LP	LP	K	LP	LP	K	LS. K	K	K
	3	P	K	LP	LP	LP	LP. K	P	LS. K	LP	LP	K	LS. K	K	K
	4	P	K	LP	LP	LP	LP. K	P	LS. K	LP	LP. K	K	LS. K	K	LS. K
	5	P	K	LP	LP	LP	LP. LS. K	K	LS. K	K	LP. LS. K	K	LS. K	P	LS. K
	6	P	K	LP	LP	LP	LP. LS	K	LS. K	LP	LP. LS. K	K	LS. K	P	LS. K
	7	P	K	LP	LP	K	LP. LS	K	LS. K	LP	LP. LS	K	LS. K	P	LS. K
	8	P	K	LP	LP	LP	LP. LS. K	P	LS. K	LP	LP. LS. K	K	LS. K	P	LS. K

T = 250															
Break location															
Break magnitude		0.2								0.5				0.8	
		0.3		0.5		0.7		0.9		0.7		0.9		0.9	
2	1	K	P	LP	P	LS. LP	K	LS	K	LP	LP	LP. LS	K	K	K
	2	LP	BP	LS	LS	LS. LP	LP	LP	LP	BP	BP	BP	LS	BP	LS
	3	LP	P	LP	LP	LP	LP	LP	K	LP	K	P	K	K	P
	4	LS	P	LP	LP	LP	LP. K	LS	K	LP	LP. K	LP	K	K	LS
	5	P	LS	LP	LP	LP	LP. LS. K	LP	LS. K	LP	LP. LS. K	P	K	K	LS
	6	P	LS	LP	LP	LP	LP. LS. K	LS	LS. K	LP	LP. LS. K	K	LS. K	K	LS. K
	7	P	LS	LP	LP. LS	LP	LP. LS K	K	LS. K	LP	LP. LS. K	K	LS. K	K	LS. K
	8	P	LS	LP	LP. LS	LP	LP. LS	K	LS. K	LP	LP. LS. K	K	LS. K	K	LS. K

T = 250															
Break location															
Break magnitude		0.2								0.5				0.8	
		0.3		0.5		0.7		0.9		0.7		0.9		0.9	
3	1	LP	P	LP	LP	LP. LS	LP. LS	LP. LS	K	LP. LS. K	LP	LP. LS. K	K	K. LS	LS
	2	LP	P	LP	K	LP	K	LP. LS	K	LP. LS. K	LP	LP. LS	K	LS	K
	3	LP	BP	LP	LS	LP. LS	P	LP. LS	LP	LP. LS	LP	LP	LP	K	BP
	4	LP	P	LP	K	LP. LS	LP. K	LP. LS	K	LP	LP	LP. K	K	K	K
	5	LS	P	LP	LP	LP	LP	LP. LS	K	LP. LS	LP	LP		LS	LS
	6	LP	K	LP	LP	LP	LP. LS. K	LP. LS	LS	LP. LS	LP	LP	LS	LS	LS
	7	LS	K	LP	LP	LP	LP. LS. K	LP. LS	LS	LP. LS. K	LP. LS. K	LP. LS. K	LS. K	LS	LS
	8	LS	K	LP. LS	LP	LP. LS	LP. LS. K	LS	LS. K	LP. LS. K	LP. LS. K	LP. LS. K	LS. K	K	LS. K

T = 250															
Break location															
Break magnitude		0.2								0.5				0.8	
		0.3		0.5		0.7		0.9		0.7		0.9		0.9	
4	1	K	P	LP	LP	LP. LS	LP. LS	LP. LS	K	LP. LS	LP	LP. LS. K	LS	K. LS	LS
	2	LS	P	LP	LP	LP. LS	LP. K	LP. LS	K	LP. LS	LP	LP. LS. K	LS. K	K. LS	LS
	3	LP. LS	P	LP	K	LP. LS	LP	LP. LS	K	LP. LS. K	LP	K	K	K. P	P
	4	LP	BP	LP	P	LP. LS	LP	LP. LS	LP	BP	LS	BP	K	K	BP
	5	LP. LS	P	LP	LP	LP. LS	LP	LP. LS	K	LP	LP	K	P	K. P	LS
	6	LP. LS	P	LP	LP	LP. LS	LP. K	LP. LS	LS	LP. LS	LP	LP	LS	LS	LS
	7	LP. LS	P	LP	LP	LP. LS	LP. LS. K	LP. LS	LS. K	LP. LS	LP. LS	LP	LS	LS	LS
	8	LP. LS	LS	LP	LP	LP. LS	LP. LS. K	LP. LS	LS	LP. LS	LP. LS	LP. LS	LS. K	LS	LS

T = 250															
Break location															
Break magnitude		0.2								0.5				0.8	
		0.3		0.5		0.7		0.9		0.7		0.9		0.9	
5	1	LS	LS	LP. LS	LP	LP. LS	LP. LS	LP. LS	LS	LP. LS	LP	LP. LS	LS	LS	LS
	2	LS	LS	LP. LS	LP	LP. LS	LP. LS	LP. LS	LS. K	LP. LS. K	LP	LP. LS	LS	LS. K	LS
	3	LP. LS	P	LP. LS	LP	LP. LS	LP. LS	LP. LS		LP. LS. K	LP	LP. LS	LS	LS. K	LS
	4	LP	P	LP. LS	P	LP. LS	K. P	LP. LS	K	LP. LS. K	LP	LP. K	K	K	P
	5	LP	BP	LP	LP	LP. LS	P	LP. LS	LP	K	LP	BP	LP	BP	K
	6	LP. LS	P	LP. LS	LP. K	LP. LS	K. P	LP. LS	P	LP. K	LP	LP. K	K	LS	LS
	7	LP. LS	P	LP. LS	LP	LP. LS	LP	LP. LS	K	LP	LP	LP. LS. K	LS	LS	LS
	8	LP. LS	P	LP. LS	LP	LP. LS	LP. LS	LP. LS	LS	LP. LS	LP	LP. LS	LS	LS	LS

T = 500															
Break location															
Break magnitude		0.2								0.5				0.8	
		0.3		0.5		0.7		0.9		0.7		0.9		0.9	
1	1	LP	BP	LP	LS	LP	LS	LP	P	K	K	P	K	BP	K
	2	P	P	LP	K	LP	K	LP	K	LP	LP	LP	K	K	P
	3	P	K	LP	LP. K	K	LP	LP	LS	LP	LP	K	LS. K	K	LS. K
	4	P	K	LP	LP	LP	LP. LS. K	K	LS. K	LP	LP. K	K	LS. K	P	LS. K
	5	P	K	LP	LP. K	LP	LP. LS. K	P	LS. K	LP	LP. K	K	LS. K	P	LS. K
	6	P	K. P	LP	LP. LS. K	K	LP. LS. K	K	LS. K	LP. K	LP. K	K. P	LS. K	K	LS. K
	7	P	K	LP	LP	LP	LP. LS. K	K	LS. K	LP. K	LP. LS. K	K	LS. K	K	LS. K
	8	P	K	LP	LP. K	LP	LP. LS. K	P	LS. K	LP. K	LP. LS. K	K	LS. K	K	LS. K

T = 500															
Break location															
Break magnitude		0.2								0.5				0.8	
		0.3		0.5		0.7		0.9		0.7		0.9		0.9	
2	1	LP. LS	P	LP	K	LP	LP	LP. LS	K	LP	LP. K	LP. K	K	K	P
	2	LP	K. P	LP. LS	LP	LP	P	LP	LP	BP	P	BP	LP	BP	BP
	3	LP	K	LP	K	LP	LP	LP. LS	K	LP	K	LP. K	K	K	K
	4	LS	K	LP	LP. K	LP. LS	LP. K	LP. LS	P. K	LP	K	K	P. K	LS	P
	5	LS	K	LP	LP. LS. K	LP	LP. K	K	LS. K	LP. K	LP. LS. K	LP. K	LS. K	P. K	LS. K
	6	P	K	LP. LS	LP. LS. K	LP	LP. LS. K	K	LS. K	LP. K	LP. LS. K	K. P	LS. K	K	LS. K
	7	P	K	LP	LP. LS. K	LP	LP. L. KS	K	LS. K	LP. K	LP. LS. K	LP	LS. K	K	LS. K
	8	P	K	LP	LP. LS. K	LP	LP. LS. K	K	LS. K	LP. K	LP. LS. K	P	LS. K	P. K	LS. K

T = 500															
Break location															
Break magnitude		0.2								0.5				0.8	
		0.3		0.5		0.7		0.9		0.7		0.9		0.9	
3	1	LP. LS	LS	LP. LS	LP	LP. LS	LP	LP. LS	K	LP. K	LP. K	LS. K	K. P	K	K. P
	2	LS	LS	LP. LS	LP	LP. LS	LP	LP. LS	K	LP. LS	P	LP. K	P	K	P
	3	LP. LS	BP	LP. LS	LS	LP. LS	LS	LP. LS	P	LP	LP	K	LP	K	BP
	4	LP. LS	P	LP. LS	LP	LP. LS	LP	LP. LS	K	LP. K	K	LP. LS. K	K	K. P	P
	5	LP. LS	LS	LP. LS	LP	LP. LS	LP	LS	K. P	LP	LP. K	LP	LS	LS	LS
	6	LS	LS. K	LP. LS	LP. K	LP	LP. K	LS	LS	LP. K	LP. LS. K	LP	LS	LS	LS
	7	LS	LS. K	LP	LP. K	LP. LS	LP. LS. K	LS	LS	LP. K	LP. LS. K	LP. K	LS. K	LS	LS
	8	LS	LS. K	LP	LP. LS. K	LP. LS	LP. LS. K	LS	LS. K	LP. K	LP. LS. K	LP. K	LS. K	LS. K	LS. K

T = 500															
Break location															
Break magnitude		0.2								0.5				0.8	
		0.3		0.5		0.7		0.9		0.7		0.9		0.9	
4	1	LS	LS	LP. LS. K	LP. LS	LP. LS. K	LP. LS	LP. LS	LS	LP. LS. K	LP. LS	LP. LS. K	LP. K	LS. K	LS
	2	LP. LS. K	LS	LP. LS	LP. LS	LP. LS	LP. LS. K	LP. LS	LS	LP. LS. K	LP	LP. LS. K	LP	LS. K	LS. K
	3	LP. LS	P	LP. LS	LP	LP. LS	LP	LP. LS	K	LP. K	LP. K	LP. LS. K	LP	K	K
	4	LP. LS	BP	LP. LS	LP	LP	P	LP. LS	LS	BP	LS	K	LP	BP	K
	5	LP. LS	K	LP. LS	K	LP. LS	K	LP. LS	K	LP	LP	LP. LS	LP	K. P	P
	6	LP. LS	K. P	LP. LS	LP	LP. LS	LP. K	LP. LS	K	LP. LS	LP	LP. LS. K	LP	LS	LS. K
	7	LP. LS	LS	LP. LS	LP. LS	LP. LS	LP. LS	LP. LS	LS	LP. LS	LP. LS	LP. LS	LP. LS. K	LS	LS
	8	LS	LS. K	LP. LS	LP. K	LP. LS	LP. LS. K	LP. LS	LS. K	LP. LS	LP. LS	LP. LS	LS	LS	LS

T = 500															
Break location															
Break magnitude		0.2								0.5				0.8	
		0.3		0.5		0.7		0.9		0.7		0.9		0.9	
5	1	LS	LS	LP. LS. K	LP. LS	LP. LS	LP. LS	LP. LS	LP. K	LP. LS. K	LP. LS	LP. LS. K	LS	LS. K	LS
	2	LP. LS. K	LS	LP. LS. K	LP. LS. K	LP. LS	LP. LS. K	LP. LS	LP. K	LP. LS. K	LP. LS	LP. LS. K	K. P	LS. K	LS
	3	LP. LS	K. P	LP. LS. K	LP. LS	LP. LS	LP	LP. LS	K. P	LP. LS	LP. LS	LP. LS. K	K	LS. K	K
	4	LP. LS	P	LP. LS	K	LP. LS	LP	LP. LS	K	LP. K	P	LP. K	K	K	P
	5	LP. LS	BP	LP. LS	LS	LP. LS	LS	LP	LP	BP	K	BP	LS	K	LS
	6	LP. LS	P	LP. LS	K	LP. LS	K	LP. LS	K	LP	LP	LP. LS	K	K. P	K
	7	LP. LS	K	LP. LS	LP. K	LP. LS	LP. K	LP. LS	K	LP. LS	LP	LP. LS. K	P	LS	LS. K
	8	LP. LS	LS	LP. LS	LP. K	LP. LS	LP. LS	LP. LS	LP. K	LP. LS	LP. LS	LP. LS	LS. K	LS	LS

*K: Kapetanios (2005) and P: Perron (2006)

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). The authors have not disclosed any funding.

Data availability The 90-day bank accepted bill rate of Australia can be gathered from https://www.oecd.org/sdd/oecdmaineconomicindicatorsmei.htm, accessed on 3 February 2024. The 3-month treasury bill rate of the US and inflation rate of the US and Australia data can be obtained from https://data.imf.org/?sk=4c514d48-b6ba-49ed-8ab9-52b0c1a0179b, accessed on 4 February 2024.

Declarations

Conflict of interest No potential conflict of interest was reported by the author(s).

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica, 59*(3), 817–858. https://doi.org/10.2307/2938229

Andrews, D. W. K., & Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica, 62*(6), 1383. https://doi.org/10.2307/2951753

Baharumshah, A. Z., Chan, T., & Fountas, S. (2005). A panel study on real interest rate parity in East Asian Countries: Pre- and post-liberalization era. *Global Finance Journal, 16*(1), 69–85. https://doi.org/10.1016/j.gfj.2005.05.005

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica, 66*(1), 47. https://doi.org/10.2307/2998540

Bai, J., & Perron, P. (2003a). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics, 18*(1), 1–22. https://doi.org/10.1002/jae.659

Bai, J., & Perron, P. (2003b). Critical values for multiple structural change tests. *The Econometrics Journal, 6*(1), 72–78. https://doi.org/10.1111/1368-423X.00102

Banerjee, A. (1992). Recursive and sequential tests of the unit-root and trend-break hypotheses: Theory and international evidence. *Journal of Business & Economic Statistics, 80*(3), 271–287.

Basawa, I. V., Mallik, A. K., McCormick, W. P., Reeves, J. H., & Taylor, R. L. (1991). Bootstrapping unstable first-order autoregressive processes. *The Annals of Statistics, 19*(2), 1098–1101.

Baxter, M., & Robert, K. G. (1993). Fiscal policy in general equilibrium. *American Economic Review, 83*(3), 315–334.

Caporale, T., & Grier, K. (2005). How smart is my dummy? Time series tests for the influence of politics. *Political Analysis, 13*(1), 77–94.

Escobar, M., Götz, B., Seco, L., & Zagst, R. (2009). Pricing A Cdo on stochastically correlated underlyings. *Quantitative Finance, 10*(3), 265–277. https://doi.org/10.1080/14697680802629418

Felmingham, B., & Mansfield, P. (2003). A note on the stability of real interest rates in Australia. *International Review of Economics & Finance, 12*(4), 517–524.

García, R., & Perron, P. (1996). An analysis of the real interest rate under regime shifts. *The Review of Economics and Statistics, 78*(1), 111. https://doi.org/10.2307/2109851

Hansen, B. E. (2001). The new econometrics of structural change: dating breaks in U.S. labour productivity. *Journal of Economic Perspectives*, *15*(4), 117–128. https://doi.org/10.1257/jep.15.4.117

Hegwood, N. D., & Papell, D. H. (1998). Quasi purchasing power parity. *International Journal of Finance & Economics, 3*(4), 279–289.

Hoffmann, M., & MacDonald, R. (2009). Real exchange rates and real interest rate differentials: A present value interpretation. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1365251
Hong, H., Bian, Z., Chen, N., & Su, C. (2020). Does interest rate liberalisation affect the constancy of mean interest rates in China? *Journal of Financial Regulation and Compliance, 28*(4), 555–568. https://doi.org/10.1108/JFRC-10-2019-0129
Horowitz, J. L. (2001). The bootstrap. In J. J. Heckman & E. Leamer (Eds.), *Handbook of econometrics* (Vol. 5, pp. 3159–3228). Elsevier.
Jostova, G., & Philipov, A. (2005). Bayesian analysis of stochastic betas. *Journal of Financial and Quantitative Analysis, 40*(4), 747–778. https://doi.org/10.1017/s0022109000001964
Kapetanios, G. (2005). Unit-root testing against the alternative hypothesis of up to m structural breaks. *Journal of Time Series Analysis, 26*(1), 123–133. https://doi.org/10.1111/j.1467-9892.2005.00393.x
Lee, J., & Strazicich, M. C. (2003). Minimum lagrange multiplier unit root test with two structural breaks. *The Review of Economics and Statistics, 85*(4), 1082–1089. https://doi.org/10.1162/003465303772815961
Lumsdaine, R. L., & Papell, D. H. (1997). Multiple trend breaks and the unit-root hypothesis. *The Review of Economics and Statistics, 79*(2), 212–218. https://doi.org/10.1162/003465397556791
Mishra, A., Moosa, I., Tawadros, G., & Mishra, V. (2023). The effect of political and bureaucratic regime changes on Australia's real interest rate. *International Review of Economics & Finance, 85*, 124–136. https://doi.org/10.1016/j.iref.2023.01.016
Namba, A. (2017). Bootstrap test for a structural break under possible heteroscedasticity. *Communications in Statistics-Simulation and Computation, 46*(5), 4127–4139. https://doi.org/10.1080/03610918.2015.1107581
Narayan, P. K., & Popp, S. (2010). A new unit root test with two structural breaks in level and slope at unknown time. *Journal of Applied Statistics, 37*(9), 1425–1438. https://doi.org/10.1080/02664760903039883
Neely, C. J., & Rapach, D. E. (2008). Real interest rate persistence: evidence and implications. *Federal Reserve Bank of St. Louis Review*, *90*(6), 609–641.
Nelson, C. R., & Plosser, C. I. (1982). Trends and random walks in macroeconmic time series: Some evidence and implications. *Journal of Monetary Economics, 10*(2), 139–162.
Newey, W. K., & West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 777–787.
Nunes, L. C., Newbold, P., & Kuan, C. M. (1997). Testing for unit roots with breaks: Evidence on the great crash and the unit root hypothesis reconsidered. *Oxford Bulletin of Economics and Statistics, 59*(4), 435–448.
Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica, 57*(6), 1361. https://doi.org/10.2307/1913712
Perron, P. (1990). Testing for a unit root in a time series with a changing mean. *Journal of Business & Economic Statistics, 8*(2), 153–162.
Perron, P. (1997). Further evidence on breaking trend functions in macroeconomic variables. *Journal of Econometrics, 80*(2), 355–385.
Perron, P. (2006). Dealing with structural breaks. *Palgrave Handbook of Econometrics, 1*, 278–352.
Rapach, D. E., & Wohar, M. E. (2005). Regime changes in international real interest rates: Are they a monetary phenomenon? *Journal of Money, Credit, and Banking, 37*(5), 887–906. https://doi.org/10.1353/mcb.2005.0057
Rappoport, P., & Reichlin, L. (1989). Segmented trends and non-stationary time series. *The Economic Journal, 99*(395), 168–177. https://doi.org/10.2307/2234078
Schmidt, P., & Phillips, P. (1992). Lm tests for a unit root in the presence of deterministic trends. *Oxford Bulletin of Economics and Statistics, 54*, 257–287. https://doi.org/10.1111/j.1468-0084.1992.tb00002.x
Sen, P. K., & Singer, J. M. (1993). Stochastic convergence. *Large Sample Methods in Statistics*. https://doi.org/10.1007/978-1-4899-4491-7_2
Smallwood, A. D., & Norrbin, S. C. (2007). An encompassing test of real interest rate equalization*. *Review of International Economics, 16*(1), 114–126. https://doi.org/10.1111/j.1467-9396.2007.00731.x
Stine, R. A. (1987). Estimating properties of autoregressive forecasts. *Journal of the American Statistical Association, 82*(400), 1072–1078. https://doi.org/10.2307/2289383

Taylor, M. P., Peel, D. A., & Sarno, L. (2001). Nonlinear mean-reversion in real exchange rates: Toward a solution to the purchasing power parity puzzles. *International Economic Review, 42*(4), 1015–1042.

Uçar, N., & Omay, T. (2009). Testing for unit root in nonlinear heterogeneous panels. *Economics Letters*, *104*(1), 5–8.

Zivot, E., & Andrews, D. W. K. (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business & Economic Statistics, 20*(1), 25–44. https://doi.org/10.1198/073500102753410372

Authors and Affiliations

Özge Çamalan[1] · Esra Hasdemir[2] · Tolga Omay[1] · Mustafa Can Küçüker[1]

✉ Tolga Omay
omay.tolga@gmail.com

Özge Çamalan
ozge.camalan@atilim.edu.tr

Esra Hasdemir
esra.hasdemir@atilim.edu.tr

Mustafa Can Küçüker
mustafa.kucuker@atilim.edu.tr

1 Department of Economics, Atilim University, Ankara, Turkey

2 Department of International Trade and Logistics, Atilim University, Ankara, Turkey