Marginal Likelihood Based LM Unit Root Tests
Allowing Multiple Level Shifts Under both the
Null and Alternative Hypotheses

Yujin Oh
Leeds School of Business, University of Colorado at Boulder, USA
E-mail address: ouj92@hotmail.com

Abstract

In this paper we propose new unit root tests when a time series has multiple level shifts. The proposed tests are Lagrangian multiplier type tests based on the marginal likelihood of the residuals. The marginal likelihood is free from the nuisance mean parameters, which are integrated out, and it is a fort from an aspect of handling the nuisance parameters. The limiting null distributions of the proposed tests are the $\chi^2$ distributions, and are affected not by the size and the location of breaks but only by the number of breaks.

We set the level shifts under both the null and the alternative hypotheses to relieve a possible vagueness in interpreting test results in empirical work. The null hypothesis implies a unit root process with level shifts and the alternative connotes a stationary process with level shifts. The Monte Carlo simulation shows that our tests are locally more powerful than the OLSE based tests, and that the powers of our tests, in a fixed time span, remain stable regardless the number of breaks. In our application, we employ the data which are analyzed by Perron (1990), and some results differ from Perron’s (1990).

JEL classification: C12, C15, and C22

Key words: Unit root; Tests; Time Series; Level Shifts; Marginal likelihood
1. Introduction

Since Perron (1989) shows in his seminal work that in the presence of a structural break the standard Augmented Dickey-Fuller (hereafter, ADF) tests are biased towards the non-rejection of the null hypothesis, a number of papers have developed unit root tests allowing for structural breaks. And it is a commonly documented finding that unit root tests in finite samples lose power in the presence of level shifts in the literature. The unit root tests already suffer low power, and neglecting the level shifts worsen the asymptotic local power of the ADF type tests. On this account, considering structural breaks in developing unit root tests is now a point of issue.

At first the early studies set out to consider one structural break. Perron (1989, 1990) and Amsler and Lee (1995) regard the break date as known, an exogenous breakpoint, while Banerjee, Lumsdaine, and Stock (1992), Zivot and Andrews (1992), and Perron and Vogelsang (1992) treat the break date as unknown, an endogenous breakpoint. Allowing only one break is restrictive to apply in practice, and therefore unit root tests with two breaks have been developed such as Lumsdaine and Papell (1997), Clemente, Montanes, and Reyes (1998), Ohara (1999), Papell and Prodan (2003), and Lee and Strazicich (2003). However, fixing the number of breaks to two is still restrictive as one, and then recently unit root tests with multiple breaks have been proposed. Kapetanios (2005) extends part of the results of Banerjee et al. (1992) and Zivot and Andrews (1992) to multiple breaks. Cavaliere and Georgiev (2007) remove the level shifts from the original time series and then apply the standard ADF tests to the "dejumped" time series. Cavaliere and Georgiev consider a rather general autoregressive data generating process with additive level shifts having more general features.

Arguably the most popular unit root test with single break is the minimum test of Zivot and Andrews (1992). The endogenous breakpoint is determined by the point which has the minimum value of test statistics using
the grid searching. The test has no break under the null hypothesis and also subsequent papers, which extend the result of Zivot and Andrews (1992) to multiple breaks such as Lumsdaine and Papell (1997) and Kapetanios (2005), have no breaks under the null. Lee and Strazicich (2003) point out that rejection of the null without breaks does not necessarily imply rejection of a unit root \textit{per se}, but may imply rejection of a unit root without breaks. Similarly, the alternative does not necessarily imply trend stationarity with breaks, but may indicate a unit root with breaks. This outcome calls for a careful interpretation of test results in empirical work.

Considering the limitations noted above, we propose in this paper Lagrange multiplier (LM) type unit root tests with multiple level shifts in which the alternative hypothesis unambiguously implies a stationary time series with breaks. Our LM tests are based on the marginal likelihood of the residual vector which is free from the nuisance mean parameters. The marginal likelihood is obtained by integrating out the nuisance parameters and this is a main difference from the profile likelihood. The asymptotics of the test statistics under the null hypothesis are the $\chi^2$ distributions which differ in the number of breaks. Both the size and the location of the breakpoints do not affect the distributions. This is a nice feature and a great contrast to the others having complex distributions under the null hypothesis.

We apply our test to the ex post real-interest rate, the U.S. unemployment rate, and the Grilli-Yang commodity price index which are already examined in Perron (1990). To compare the results of Perron (1990), we apply the proposed unit root test with one level shift using the same break date. The empirical results provide an evidence that the real-interest rate is stationary process with a break but both the U.S. unemployment rate and the Grilli-Yang commodity price index are nonstationary process with a break.

The remainder of this paper is organized as follows. Section 2 proposes the test statistics and investigates the asymptotic null distributions of the tests. Section 3 provides the Monte Carlo simulation results. Section 4
presents our empirical application. Section 5 concludes and all proofs are given in the Appendixes.

2. Test Statistics

1) Single break

We consider the following model

\[ y_t = \mu_1 + \delta D_t + u_t, \quad t = 1, \ldots, T \]

\[ u_t = \rho u_{t-1} + \epsilon_t, \]  

where \( D_t = 1 \) if \( t > T_B \) and \( D_t = 0 \) otherwise, and \( T_B \) stands for the break point. We assume that \( T_B = \lambda T \) is given, \( 0 < \lambda < 1 \), with both \( T \) and \( T_B \) integer-valued. Perron (1990) refers to \( \lambda \) as the break fraction. We let \( \mu_2 = \mu_1 + \delta \), then the \( \mu_1 \) and the \( \mu_2 \) respectively denote the mean for the first and the second regime. The \( \epsilon_t \) is the error term that is independent and normally distributed with mean 0 and variance \( \sigma^2 \).

We are interested in testing the null hypothesis \( H_0 : \rho = 1 \) versus the alternative \( H_1 : |\rho| < 1 \). Since the model includes the break under both the null and the alternative hypotheses, our tests give more clear explanation for empirical results than do other tests which do not allow breaks under the null hypothesis. If the break is not hold under the null hypothesis, then the null may implicate a unit root \textit{per se} or a unit root without break. And the alternative may imply stationarity with break or a unit root with break. However, in our tests, the null hypothesis implies that the time series is nonstationary with one shift occurring in its mean level, while the alternative hypothesis means that the time series is stationary except for this change in the mean.
Now we develop an LM test based on the marginal likelihood which is free from the nuisance parameter $\mu_1$ and $\mu_2$. To begin, we construct the full likelihood, and then develop the marginal likelihood. The full likelihood is presented as follows

$$L(\rho, \sigma^2 | \mu_1, \mu_2) = \prod_{i=1}^{2} \frac{|V^{-1}|^{1/2}}{(\sqrt{2\pi})^{T_i}} \exp \left[ -\frac{1}{2} (Y_i - \mu_i X_i)' V^{-1} (Y_i - \mu_i X_i) \right]$$

where $|\cdot|$ is the determinant of a matrix, $Y_1 = [y_1, y_2, \ldots, y_{T_B}]'$, $Y_2 = [y_{T_B+1}, y_{T_B+2}, \ldots, y_T]'$, $T_1 = T_B$, $T_2 = T - T_B$, the $T_i \times 1$ vector $X_i = 1_{T_i}$ $= [1, 1, \ldots, 1]'$ for $i = 1, 2$. We have the vector $Y_i - \mu_i X_i \sim N_T(0, V(\theta))$ where $\theta = (\rho, \sigma^2)$. The $k$-th diagonal elements of $V(\theta)$ are $V_{k,k} = \sigma^2/(1 - \rho^2)$, and the $(k, s)$ off-diagonal elements of $V(\theta)$ are $V_{k,s} = \sigma^2 \rho^{|k-s|}/(1 - \rho^2)$, for $k \neq s$, $k, s = 1, \ldots, T_i$. We use $V$ in lieu of $V(\theta)$ as a matter of convenience.

The following is the marginal likelihood, $L_M(\rho, \sigma^2)$, which is obtained by integrating out the nuisance parameters, $\mu_1$ and $\mu_2$, from the full likelihood.

$$L_M(\rho, \sigma^2) = \int \int L(\rho, \sigma^2 | \mu_1, \mu_2) \, d\mu_1 d\mu_2 = \int L_1 d\mu_1 \int L_2 d\mu_2$$

$$= \prod_{i=1}^{2} \frac{|V^{-1}|^{1/2}}{(\sqrt{2\pi})^{T_i-1}|X_i'V^{-1}X_i|^{1/2}} \exp \left[ -\frac{1}{2} (Y_i - \hat{\mu}_i X_i)' V^{-1} (Y_i - \hat{\mu}_i X_i) \right]$$

where $L_i$ is the full likelihood of $Y_i$, and $\hat{\mu}_i = (X_i'V^{-1}X_i)^{-1}(X_i'V^{-1}Y_i)$ which is identical to the generalized least squares estimate (GLSE) of $\mu$. We note that both $\hat{\mu}_i$ and $|V^{-1}|/|X_i'V^{-1}X_i|$, $i = 1, 2$ are well-defined at $\rho = 1$ in formula (3).

On the basis of the marginal likelihood, we can develop an LM test for the null hypothesis of nonstationary time series, $H_0 : \rho = 1$. We first establish a
test statistic for each regime. We then construct the proposed test statistic by pooling each test statistic. To build our test statistic, we now derive the score vector and the Fisher information matrix. Let $S^{(i)}(\theta)$ be the score vector for the $i$-th regime where

$$S^{(i)}(\theta) = \left[ \frac{\partial l^{(i)}_M}{\partial \rho}, \frac{\partial l^{(i)}_M}{\partial \sigma^2} \right]'$$

and let $l^{(i)}_M = \ln L^{(i)}_M(\rho, \sigma^2)$ be the $i$-th marginal log likelihood. Let $J^{(i)}$ be the inverse of the Fisher information matrix for the $i$-th regime where $J^{(i)}_{mn}$ represents the $(m, n)$-th components of $J^{(i)}$. In this context, our LM test statistic, $t_M$, from the score vector and the inverse of Fisher information matrix based on the marginal likelihood, is represented as follows.

$$t_M = \frac{1}{\sqrt{2}} \sum_{i=1}^{2} \tilde{S}^{(i)}_1 \sqrt{J^{(i)}_{11}}$$

(4)

where $S^{(i)}_1 = S^{(i)}_1(\tilde{\theta}_i)$, in which $\tilde{\theta}_i = (1, \tilde{\sigma}^2_i)$ and $\tilde{\sigma}^2_i = \frac{1}{T_i - 1} \sum_{t=F_i+1}^{L_i} (y_t - y_{t-1})^2$, and $J^{(i)}_{11}$ are

$$\tilde{S}^{(i)}_1 = \frac{T_i - 1}{4} + \frac{1}{\tilde{\sigma}^2_i} \left\{ \left( \frac{y_{L_i} - y_{F_i}}{2} \right)^2 - \frac{1}{2} \sum_{t=F_i+1}^{L_i} (y_t - y_{t-1})^2 \right\}$$

(5)

and

$$J^{(i)}_{11} = \frac{8}{(T_i - 1)(T_i - 2)}.$$ 

(6)

Here $y_{F_i}$ and $y_{L_i}$ denote the first and last observation in each regime, respectively. For the first regime, $y_{F_i} = y_1$ and $y_{L_i} = y_{T_B}$ and for the second regime, $y_{F_i} = y_1$ and $y_{L_i} = y_{T_B}$. In the following theorem, we derive the null distribution of the proposed score test statistic.
Theorem 1 Consider model (1) and let $\frac{\hat{\mu}}{\lambda} \rightarrow \lambda$ as $T \rightarrow \infty$. Then under the null hypothesis, $H_0 : \rho = 1$, we have $t_M \xrightarrow{d} \frac{1}{2}\{\chi^2(2) - 2\}$ as $T \rightarrow \infty$.

Here $\xrightarrow{d}$ stands for the convergence in distribution, and $\chi^2(df)$ denotes the $\chi^2$ distribution with the $df$ degree of freedom. We note that the limiting null distribution is the standardized $\chi^2$ distribution and also note that the null distribution is not affected by $\lambda$.

Table 1 presents selected percentage points that permit hypothesis testing. The critical values of $T_M$ are obtained directly from the $\chi^2(2)$ distribution.

< Here table 1 >

The critical values, for a given size level, are relatively smaller in absolute value than those of Perron (1990) in the left tail of the distribution. Therefore we may expect that our test can have better power performance.

2) Multiple breaks

We now consider $K$ multiple breaks in level and the model is

$$y_t = \mu_1 + \sum_{k=1}^{K} \delta_k D_t(T_k) + u_t,$$

$$u_t = \rho u_{t-1} + \epsilon_t,$$  \hspace{1cm} (7)

for $t = 1, \ldots, T$, where $T_1, \ldots, T_K$ denote the given break points and $D_t(T_k) = 1$ if $t > T_k$ and $D_t(T_k) = 0$ otherwise. We assume that $T_1 < T_2 < \cdots < T_K$. The $\mu_1$ denotes the overall mean and $\epsilon_t$ is the error term that is independent and normally distributed with mean 0 and variance $\sigma^2$. We are interested
in testing the null hypothesis $H_0 : \rho = 1$ versus the alternative $H_1 : |\rho| < 1$. The null hypothesis also means that the time series is nonstationary with more than one shift occurring in its mean level, while the alternative hypothesis purports that the time series is stationary except for changes in the mean. If we let $\mu_k = \mu_1 + \sum_{i=1}^{k-1} \delta_i$, $k = 2, ..., K + 1$, then the full likelihood is presented as follows

$$L(\rho, \sigma^2 | \mu_1, ..., \mu_{K+1}) =$$

$$\prod_{k=1}^{K+1} \frac{|V^{-1}|^{1/2}}{(\sqrt{2\pi})^{n_k}} \exp \left[ -\frac{1}{2} (Y_k - \mu_k X_k)' V^{-1} (Y_k - \mu_k X_k) \right]$$

where $n_k = T_k - T_{k-1}$, $T_0 = 0$, $T_{K+1} = T$, $Y_{K+1} = [y_{T_k+1}, y_{T_k+2}, ..., y_{T_{k+1}}]'$, and the $n_k \times 1$ vectors $X_k = 1_{n_k} = [1, 1, ..., 1]'$. Let us now formulate the marginal likelihood, $L_M(\rho, \sigma^2)$, which is defined as

$$L_M(\rho, \sigma^2) = \int \cdots \int L(\rho, \sigma^2 | \mu_1, ..., \mu_{K+1}) \, d\mu_1 \cdots d\mu_{K+1}$$

$$= \prod_{k=1}^{K+1} \frac{|V^{-1}|^{1/2}}{(\sqrt{2\pi})^{n_k-1}|X_k'V^{-1}X_k|^{1/2}} \exp \left[ -\frac{1}{2} (Y_k - \hat{\mu}_k X_k)' V^{-1} (Y_k - \hat{\mu}_k X_k) \right]$$

where $\hat{\mu}_k = (X_k'V^{-1}X_k)^{-1}(X_k'V^{-1}Y_k)$. In the same context, the test statistic is given by

$$t_M = \frac{1}{\sqrt{K+1}} \sum_{k=1}^{K+1} S_1^{(k)} \sqrt{J_{11}^{(k)}}$$

where
\[ S_1^{(k)} = S_1^{(k)}(\hat{\theta}_k) = \frac{n_k - 1}{4} + \frac{1}{\tilde{\sigma}_k^2} \left\{ \left( \frac{y_{L_k} - y_{F_k}}{2} \right)^2 - \frac{1}{2} \sum_{t=F_k+1}^{L_k} (y_t - y_{t-1})^2 \right\} \] (11)

and

\[ J_{11}^{(k)} = \frac{8}{(n_k - 1)(n_k - 2)} . \] (12)

Here \( y_{F_k} \) and \( y_{L_k} \) denote the first and last observation in each regime, respectively, and \( \tilde{\theta}_k = (1, \tilde{\sigma}_k^2) \) and \( \tilde{\sigma}_k^2 = \frac{1}{n_k - 1} \sum_{t=F_k+1}^{L_k} (y_t - y_{t-1})^2 \). In the following theorem, we derive the null distribution of the proposed score test statistic. We note that the null distribution is not affected by \( \lambda \).

**Theorem 2** Consider model (7) with \( K \) breaks and let \( n_k \to \infty \) for all \( k \) as \( T \to \infty \). Then under the null hypothesis, \( H_0 : \rho = 1 \), we have \( t_M \xrightarrow{d} - \frac{1}{\sqrt{2(K+1)}} \{ \chi^2(K + 1) - (K + 1) \} \).

Here \( \xrightarrow{d} \) stands for the convergence in distribution, and \( \chi^2(df) \) denotes the \( \chi^2 \) distribution with \( df \) degree of freedom. The distribution depends on the number of breaks but free from the location and size of the breaks. We note that the limiting null distribution is the standardized \( \chi^2 \) distribution.

Table 2 presents selected percentage points that permit hypothesis testing. The critical values of \( t_M \) are obtained directly from the \( \chi^2(\cdot) \) distribution.

< Here table 2 >

9
The critical values, for a given size of the test, are relatively smaller (in absolute value) than other tests’ critical values in the left tail of the distribution. Therefore one would expect loss less power when the alternative is a stationary process.

3. Simulations

1) Single break

In this section, we present the results of the Monte-Carlo experiments to investigate finite sample performance of the proposed tests. For the case of single break, we compare our tests with the Perron (1990) tests. The experiments examine power performances of the two unit root tests for the model (1) under the general setup of the null hypothesis, $H_0: \rho = 1$, and the alternative, $H_1: |\rho| < 1$.

We set $\mu_1 = 0$, $\delta \sim N(0, 1)$, $\lambda = 0.1, 0.3, 0.5$, $y_0 = 0$, and $\epsilon_t \sim N(0, 1)$. Once these parameters are generated independently of $\epsilon_t$, they fixed throughout replication and we replicate 10,000 times. Table 3 shows the size and powers of the two test statistics at the 5% level with $\rho = 0.99, 0.95$ as alternatives.

<Here Table 3>

The simulation results show that the sizes of our test are more distorted than those of Perron’s (1990) and that our test is more powerful than the Perron (1990) test especially for the models for relatively short time spans.

2) Multiple breaks
We also conduct another simulation to examine the size and power performances of the proposed tests. We set $\mu_1 = 0$, $\delta_k \sim N(0, 1)$, $y_0 = 0$, and $\epsilon_t \sim N(0, 1)$. The breaks are located on every $\left\lfloor \frac{T}{K} \right\rfloor$-th observation where $\lfloor x \rfloor$ indicates the greatest integer less than or equal to $x$. This may not affect the power performance according to the table 3. We perform the simulation for the case of $T=50$, 100, and 200, $K=2, 3, 4$, and $\rho=1, 0.99, 0.95, 0.90$, and 0.80. Since the Perron (1990) tests does not allow for multiple breaks, we just summarize the results from our LM tests in table 4.

< Here table 4 >

Table 4 shows that our test has a stable size and that the size adjusted powers are robust to the number of breaks. The results document that our tests have robust powers to the location and the size of breaks. The sizes of our tests are stable for all cases and the power is relatively improved.

4. Empirical Applications

1) Data

In this section we present empirical applications involving the data which have shown breaks in the mean level in Perron (1990). Although most economic time series seem to have trends, there are, however, a number of time series without trend having level shifts in the mean. The data are the U.S. ex post real-interest rate, the U.S. unemployment rate, and the Grilli-Yang real-commodity-price index. Perron (1990) points out that a quick glance at the graph of a time series reveals the presence of a sudden change in the mean level of the series at a given time period. He concentrates on the issue of deciding whether a particular series is characterized by stationary deviations around a shifting mean function or by an integrated process in the important case in which only one shift occurs in the series.
The first data is the quarterly U.S. ex post real-interest rate over the period 1961Q1-1986Q3 which is constructed by using the rate on three-month treasury bills deflated by the consumer price index inflation rate. This series exhibits a level shift in its mean around the year 1980. Perron picks 1980Q3 as the break point and we follow this strategy. The second is the quarterly U.S. unemployment series over the period 1948Q1-1988Q3 which has attracted some attention in economic literature. The break point is chosen to be 1973Q4 like in Perron (1990) which is consistent with Evans (1989). The third is the yearly Grilli-Yang real-commodity-price index over the period 1900-1983. Cuddington and Urzua (1989) examines the series, and they also argue that the series exhibits a marked change in mean in the year 1920.

2) Results from Unit Root Tests

Table 5 presents the results of the tests. As a result of our test, the unit root hypothesis can be rejected only for the real-interest rate, but both the unemployment rate and the Grilli-Yang CPI series do not allow for rejecting the null hypothesis. We, therefore, can conclude that the interest rate is stationary with a level shift in the year 1980. However, the unemployment rate and the Grilli-Yang CPI are random walks.

Perron (1990) also applies the Dickey-Fuller test to three periods: full sample, pre-break period, and post-break period. In most cases the unit root hypothesis can not be rejected except for the second regime of the Grilli-Yang real-commodity-price index.
3) Selection of Break

The location of the break point ($T_B$) can be endogenously determined by the grid searching method as of Zivot and Andrews (1992). The method is of calculating test statistics for all possible break points and considering the date having the minimum test statistics as the break date. Thus the break date can be determined which satisfies the following formula:

$$\inf_{\lambda \in \Lambda} t_M(\lambda) = \inf_{\lambda \in \Lambda} t_M(\lambda)$$  \hspace{1cm} (13)

where $\lambda = T_B/T$ and $\Lambda$ is a closed subset of $(0,1)$. We also applied the $t_M$ test statistic for all observed data and the result is summarized in table 6.

The detected break point is quite close to the exogenous break point for each time series. Therefore, for a time series with one structural break, the break date can be endogenously selected by the proposed test. We leave more breaks for further study.

5. Concluding Remarks

In this paper, we propose LM type unit root tests for a time series with level shifts in its mean. Our tests are based on the marginal likelihood of the residual vector which is free of nuisance mean parameters. The marginal likelihood is obtained by integrating out the mean parameters. The limiting distributions of the proposed tests under the null are the $\chi^2$ distributions. The distribution is not affected by the size and the location of the breaks.
but this only depends on the total number of breaks. This is one of good features of our tests.

In the Monte Carlo study, the proposed tests show another good property under multiple breaks which the power is stable for the number of breaks. Only matter for the power is the time span. The longer a time series has observations, the more the test statistic has power. Also, in the case of single break, the proposed tests are locally more powerful than the OLSE based tests especially for relatively short time spans.

We apply our tests to real data (the quarterly U.S. ex post real-interest rate, the quarterly U.S. unemployment rate, and the yearly Grilli-Yang real-commodity-price index) which Perron (1990) examines as time series without trend having single break. The U.S. ex post real-interest rate turns out to be stationary. We also make use of the proposed tests in detecting the break point for a single break of the real data in the framework of minimum test statistic. The selected break point is almost identical to the exogenous change date in previous literatures.
Appendix A. Likelihood and its derivatives

In order to calculate score vector and the Fisher information matrix, we need to derive the first and second derivatives of $l^{(i)}_M$, $l^{(i)}_M = \ln L^{(i)}_M(\rho, \sigma^2)$. Each components of the marginal likelihood, formula (9), can be represented as the followings:

$$\hat{\mu}_i(\rho) = (X'_iV^{-1}X_i)^{-1}X'_iV^{-1}Y_i = \frac{2\rho(y_{F_i} + y_{L_i}) + (1 - \rho)T_i\bar{y}_i}{2\rho + T_i(1 - \rho)} ,$$

$$\frac{\partial \hat{\mu}_i}{\partial \rho}(\rho) = \frac{T(y_{F_i} + y_{L_i} - 2\bar{y}_i)}{(2\rho + T_i(1 - \rho))^2} ,$$

$$\hat{\mu}_i(1) = \frac{(y_{F_i} + y_{L_i})}{2} ,$$

$$\frac{\partial \hat{\mu}_i}{\partial \rho}(1) = \frac{1}{2} \left\{ \frac{T}{2} (y_{F_i} + y_{L_i}) - T_i\bar{y}_i \right\} ,$$

$$|V^{-1}| = (1 - \rho^2)/\sigma^2T_i ,$$

$$|X'_iV^{-1}X_i| = (1 - \rho)\{(T_i - 2)(1 - \rho) + 2\}/\sigma^2 ,$$

$$Y'_i(V^{-1} - V^{-1}X_i(X'_iV^{-1}X_i)^{-1}X'_iV^{-1})Y_i = \frac{1}{\sigma^2} \left[ \sum_{t=F_i+1}^{L_i} (y_t - \rho y_{t-1})^2 + (1 - \rho)A_i \right] ,$$

where $A_i = (1 + \rho)y^2_{F_i} - (T_i - (T_i - 2)\rho)\hat{\mu}_i^2$. Then we have the first derivatives

$$\frac{\partial l^{(i)}_M}{\partial \rho} = \frac{1}{2(1 + \rho)} + \frac{T_i - 2}{2(T_i - (T_i - 2)\rho)}$$

$$- \frac{1}{\sigma^2} \left[ - \sum_{t=F_i+1}^{L_i} y_{t-1}(y_t - \rho y_{t-1}) - \frac{A_i}{2} + \left( \frac{1 - \rho}{2} \right) \frac{\partial A_i}{\partial \rho} \right]$$

and

15
\[
\frac{\partial l^{(i)}}{\partial \sigma^2} = -\frac{T_i - 1}{2\sigma^2} + \frac{1}{2\sigma^4} \left[ \sum_{t=F_i+1}^{L_i} (y_t - \rho y_{t-1})^2 + (1 - \rho)A_i \right].
\]

If we let \( \tilde{\theta}_i = [1, \tilde{\sigma}_i^2]' \) and \( \tilde{\sigma}_i^2 \) maximizes the marginal likelihood under \( H_0: \rho = 1 \), then \( \tilde{\sigma}_i^2 = \frac{1}{T_i - 1} \sum_{t=F_i+1}^{L_i} (y_t - y_{t-1})^2 \). The second derivatives are obtained as
\[
\frac{\partial^2 l^{(i)}}{\partial \rho^2} = -\frac{1}{2(1 + \rho)^2} + \frac{(T_i - 2)^2}{2(2\rho + T_i(1 - \rho))^2} - \frac{1}{\sigma^2} \left[ \sum_{t=F_i+1}^{L_i} y_t^2 - (T_i - 2)\hat{\mu}_i^2 + 2\hat{\mu}_i - \frac{T_i(y_{F_i} + y_{L_i} - 2\bar{y}_i)}{T_i - (T_i - 2)\rho} + \left( \frac{1 - \rho}{2} \right) \frac{\partial^2 A_i}{\partial \rho^2} \right],
\]
\[
\frac{\partial^2 l^{(i)}}{\partial \rho \partial \sigma^2} = \frac{1}{\sigma^4} \left[ -\sum_{t=F_i+1}^{L_i} y_t(y_t - \rho y_{t-1}) - \frac{A_i}{2} + \left( \frac{1 - \rho}{2} \right) \frac{\partial A_i}{\partial \rho} \right],
\]
\[
\frac{\partial^2 l^{(i)}}{\partial \sigma^4} = T_i - 1 \frac{T_i}{2\sigma^4} - \frac{1}{\sigma^6} \left[ \sum_{t=F_i+1}^{L_i} (y_t - \rho y_{t-1})^2 + (1 - \rho)A_i \right],
\]
\[
\frac{\partial A_i}{\partial \rho} = y_{F_i} + (T_i - 2)\bar{\mu}_i^2 - 2\hat{\mu}_i - \frac{T_i(y_{F_i} + y_{L_i} - 2\bar{y}_i)}{T_i - (T_i - 2)\rho}.
\]

The expectation of the observed Fisher information is
\[
E \left[ -\frac{\partial^2 l^{(i)}}{\partial \rho^2} \right] = \frac{(T_i - 1)^2}{8}, \quad E \left[ -\frac{\partial^2 l^{(i)}}{\partial \rho \partial \sigma^2} \right] = -\left( \frac{T_i - 1}{4} \right) \frac{1}{\sigma^2}, \quad E \left[ -\frac{\partial^2 l^{(i)}}{\partial \sigma^4} \right] = \frac{T_i - 1}{2\sigma^4}.
\]
since
\[ \sum_{t=F_{i+1}}^{L_i-1} E(y_t^2) = \frac{(T_i+1)(T_i-2)}{2} \sigma^2, \]
\[ E\left(\frac{y_{F_i} + y_{L_i}}{2}\right)^2 = \frac{T_i+3}{4} \sigma^2, \]
\[ E[T_i \bar{y}_i (y_{F_i} + y_{L_i})] = \frac{T_i(T_i+3)}{2} \sigma^2, \]
\[ E[\sum_{t=F_{i+1}}^{L_i} y_t y_{t-1}] = \frac{T_i(T_i-1)}{2} \sigma^2, \]
and
\[ E\left[\sum_{t=F_{i+1}}^{L_i} (y_t - y_{t-1})^2\right] = E\left[\sum_{t=F_{i+1}}^{L_i} \epsilon_t^2\right] = (T_i - 1) \sigma^2. \]

**Appendix B. Proof of Theorem 1**

Since, for each $i$, $\hat{\sigma}_i^2/(T_i - 1) = \sigma^2 + o_p(1)$ by the law of large numbers and $(y_{L_i} - y_{F_i})/(\sigma \sqrt{T_i - 1}) \xrightarrow{d} W(1)$ by the central limit theorem, we have
\[ \hat{S}_i^{(i)} \sqrt{J_i^{(i)}} \xrightarrow{d} \{W_i(1)^2 - 1\}/\sqrt{2} \quad \text{as} \quad T_B = \lambda T \to \infty. \]
Therefore directly we have the desired result $t_M \xrightarrow{d} \frac{1}{2}(W_1(1)^2 + W_2(1)^2 - 2)$ as $T \to \infty$.

**Appendix C. Proof of Theorem 2**

We omit the proof because it is very similar to the proof in the extended case to multiple breaks of theorem 1.
References


Table 1: Percentiles of the distribution of $T_M$: Single break

<table>
<thead>
<tr>
<th></th>
<th>0.01</th>
<th>0.025</th>
<th>0.05</th>
<th>0.10</th>
<th>0.90</th>
<th>0.95</th>
<th>0.975</th>
<th>0.99</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-.99</td>
<td>-.97</td>
<td>-.95</td>
<td>-.89</td>
<td>1.31</td>
<td>2.00</td>
<td>2.69</td>
<td>3.61</td>
</tr>
</tbody>
</table>
Table 2: Percentiles of the distribution of $T_M$: Multiple breaks

<table>
<thead>
<tr>
<th>K</th>
<th>0.01</th>
<th>0.025</th>
<th>0.05</th>
<th>0.10</th>
<th>0.90</th>
<th>0.95</th>
<th>0.975</th>
<th>0.99</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-1.18</td>
<td>-1.13</td>
<td>-1.08</td>
<td>-0.99</td>
<td>1.33</td>
<td>1.96</td>
<td>2.59</td>
<td>3.40</td>
</tr>
<tr>
<td>3</td>
<td>-1.31</td>
<td>-1.24</td>
<td>-1.16</td>
<td>-1.04</td>
<td>1.34</td>
<td>1.94</td>
<td>2.52</td>
<td>3.28</td>
</tr>
<tr>
<td>4</td>
<td>-1.41</td>
<td>-1.32</td>
<td>-1.22</td>
<td>-1.07</td>
<td>1.34</td>
<td>1.92</td>
<td>2.48</td>
<td>3.19</td>
</tr>
<tr>
<td>5</td>
<td>-1.48</td>
<td>-1.37</td>
<td>-1.26</td>
<td>-1.10</td>
<td>1.34</td>
<td>1.90</td>
<td>2.44</td>
<td>3.12</td>
</tr>
<tr>
<td>6</td>
<td>-1.54</td>
<td>-1.42</td>
<td>-1.29</td>
<td>-1.12</td>
<td>1.34</td>
<td>1.89</td>
<td>2.41</td>
<td>3.07</td>
</tr>
<tr>
<td>7</td>
<td>-1.59</td>
<td>-1.46</td>
<td>-1.32</td>
<td>-1.13</td>
<td>1.34</td>
<td>1.88</td>
<td>2.38</td>
<td>3.02</td>
</tr>
<tr>
<td>8</td>
<td>-1.63</td>
<td>-1.48</td>
<td>-1.34</td>
<td>-1.14</td>
<td>1.34</td>
<td>1.87</td>
<td>2.36</td>
<td>2.99</td>
</tr>
<tr>
<td>9</td>
<td>-1.66</td>
<td>-1.51</td>
<td>-1.36</td>
<td>-1.15</td>
<td>1.34</td>
<td>1.86</td>
<td>2.34</td>
<td>2.95</td>
</tr>
<tr>
<td>10</td>
<td>-1.69</td>
<td>-1.53</td>
<td>-1.37</td>
<td>-1.16</td>
<td>1.34</td>
<td>1.85</td>
<td>2.33</td>
<td>2.93</td>
</tr>
</tbody>
</table>
Table 3: Rejection probabilities (%)

<table>
<thead>
<tr>
<th>T</th>
<th>λ</th>
<th>size</th>
<th>power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ρ=1</td>
<td>ρ=0.99</td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>tₘ</td>
<td>Perron</td>
</tr>
<tr>
<td>50</td>
<td>0.1</td>
<td>8.71</td>
<td>5.79</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>6.72</td>
<td>5.83</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>6.39</td>
<td>5.42</td>
</tr>
<tr>
<td>100</td>
<td>0.1</td>
<td>7.05</td>
<td>6.01</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>6.00</td>
<td>5.76</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>5.80</td>
<td>5.10</td>
</tr>
<tr>
<td>200</td>
<td>0.1</td>
<td>5.90</td>
<td>5.21</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>5.21</td>
<td>5.13</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>5.07</td>
<td>5.03</td>
</tr>
</tbody>
</table>

Note: Tests for $H_0: \rho = 1$ in model $y_t = \rho y_{t-1} + \delta \Delta D_t + \epsilon_t; \mu_1 = 0; y_0 = 0;$ $\delta \sim N(0,1); \epsilon_t \sim N(0,1); \rho = 1.00, 0.99, 0.95; \text{Size of tests} = 5\%; \text{Once } \delta \text{ is generated independently of } \epsilon_{it}, \text{it is fixed throughout replications; number of replications} = 10,000.$
Table 4: Rejection probabilities (%)  

<table>
<thead>
<tr>
<th>ρ</th>
<th>K= 2</th>
<th>3</th>
<th>4</th>
<th>ρ</th>
<th>K= 2</th>
<th>3</th>
<th>4</th>
<th>ρ</th>
<th>K= 2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>5.92</td>
<td>5.67</td>
<td>5.12</td>
<td>6.06</td>
<td>5.61</td>
<td>5.07</td>
<td>4.93</td>
<td>5.23</td>
<td>4.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.99</td>
<td>6.23</td>
<td>5.61</td>
<td>5.76</td>
<td>6.77</td>
<td>6.49</td>
<td>6.39</td>
<td>8.59</td>
<td>8.35</td>
<td>8.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.95</td>
<td>8.90</td>
<td>7.69</td>
<td>7.76</td>
<td>12.84</td>
<td>12.68</td>
<td>12.49</td>
<td>25.12</td>
<td>24.97</td>
<td>24.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.90</td>
<td>13.73</td>
<td>12.32</td>
<td>12.14</td>
<td>25.06</td>
<td>25.65</td>
<td>25.44</td>
<td>49.00</td>
<td>52.65</td>
<td>54.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.80</td>
<td>24.90</td>
<td>24.69</td>
<td>23.40</td>
<td>50.23</td>
<td>53.76</td>
<td>55.35</td>
<td>81.15</td>
<td>87.06</td>
<td>89.78</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Tests for $H_0 : \rho = 1$ in model $y_t = \rho y_{t-1} + \sum_{k=1}^{K} \delta_k \Delta D_t(T_k) + \epsilon_t$; $\mu_1 = 0; y_0 = 0; \delta_k \sim N(0,1),$ for $k = 1, ..., K; \epsilon_t \sim N(0,1); \rho = 1.00, 0.99, 0.95, 0.90, 0.80$; Size of tests = 5%; Once $\delta_k$ are generated independently of $\epsilon_{it}$, they are fixed throughout replications; number of replications = 10,000; The breaks are considered to be happened on every $\lfloor \frac{T}{K} \rfloor$-th observation.
Table 5: Full-sample unit root tests with a level shift

<table>
<thead>
<tr>
<th>Series</th>
<th>Period</th>
<th>T</th>
<th>$t_M$</th>
<th>$t_{\alpha^*}$</th>
<th>$t_{\hat{\alpha}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real-interest rate</td>
<td>1961Q1-1986Q3</td>
<td>103</td>
<td>-0.96*</td>
<td>-2.68</td>
<td>-3.71*</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>1948Q1-1988Q3</td>
<td>163</td>
<td>-0.66</td>
<td>-3.87*</td>
<td>-3.84*</td>
</tr>
<tr>
<td>Grilli-Yang index</td>
<td>1900-1983</td>
<td>84</td>
<td>-0.85</td>
<td>-5.25*</td>
<td>-2.46</td>
</tr>
</tbody>
</table>

NOTE: $T$ is the total number of observations, $t_M$ is the proposed test statistic, and both $t_{\alpha^*}$ and $t_{\hat{\alpha}}$ are the full-sample unit root test with changing mean of Perron (1990). The $t_{\alpha^*}$ uses additive-outlier method and the $t_{\hat{\alpha}}$ uses innovational-outlier method.
Table 6: Selection of the break

<table>
<thead>
<tr>
<th>Series</th>
<th>Period</th>
<th>T</th>
<th>(T_B)</th>
<th>(\hat{T}_B)</th>
<th>(\inf t_M(\hat{T}_B/T))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real-interest rate</td>
<td>1961Q1-1986Q3</td>
<td>103</td>
<td>79</td>
<td>78</td>
<td>-0.98*</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>1948Q1-1988Q3</td>
<td>163</td>
<td>103</td>
<td>102</td>
<td>-0.75</td>
</tr>
<tr>
<td>Grilli-Yang index</td>
<td>1900-1983</td>
<td>84</td>
<td>21</td>
<td>23</td>
<td>-0.87</td>
</tr>
</tbody>
</table>

NOTE: \(T\) is total number of observation, \(T_B\) is given break date, and \(\hat{T}_B\) is the detected break point from \(\inf t_M\).