

# Strong Rules for Detecting the Number of Breaks in a Time Series

Filippo Altissimo\* and Valentina Corradi\*\*

\* Bank of Italy

Research Department

via Nazionale, 91

Roma, 00184, Italy

tel. (39) 064792 3441

email: altissimo.filippo@insedia.interbusiness.it

\*\* Queen Mary and Westfield College

Department of Economics

London, U.K., E1 4NS

tel. (44 171) 975 5086

email: v.corradi@qmw.ac.uk

December, 1999

## Abstract

This paper proposes a new approach for detecting the number of structural breaks in a time series when estimation of the breaks is performed one at the time. We consider the case of shifts in the mean of a possibly nonlinear process, allowing for dependent and heterogeneous observations. This is accomplished through a simple, sequential, almost sure rule ensuring that, in large samples, both the probabilities of overestimating and underestimating the number of breaks are zero. A new estimator for the long run variance which is consistent also in the presence of neglected breaks is proposed. The finite sample behavior is investigated via a simulation exercise. The sequential procedure, applied to the weekly Eurodollar interest rate, detects multiple breaks over the period 1973-1995.

Keywords: functional law of the iterated logarithm, multiple structural breaks, sequential hypothesis testing.

JEL Classification: C20, C22.

---

We thank Karim Abadir, Giovanni Forchini, Patrick Marsh, Hashem Pesaran, Elias Tzavalis, Robert Taylor, Andrew Tramayne, Tim Vogelsang and the seminar participants at the European Meeting of the Econometric Society 1999, University of York, Queen Mary and Westfield College for comments and suggestions on earlier versions of this paper. We remain, however, solely responsible for any omission and mistakes, as well as for the views expressed in the paper, which do not reflect those of the Bank of Italy.

# 1 Introduction

The analysis of structural breaks in time series data has received a lot of attention over the last ten-fifteen years, both from a theoretical and an empirical point of view. There are several reasons why it is important to test for structural breaks and, if we reject the null, estimate them. Parameter stability is typically kept as a maintained assumption in classical hypothesis testing. Violation of such maintained assumption can lead to unreliable inference. A very well known case is that of testing for unit roots in the presence of broken trends. Perron in his (1989) seminal paper has shown that the failure to reject the null of unit roots may be attributable to the failure of taking into account the presence of broken trends.<sup>1</sup> On a different ground, out of sample prediction is typically performed under the maintained assumption of parameter stability (see e.g. Hendry (1997) for an overview). If we are interested in out of sample forecast accuracy, it is crucial to be able to detect and estimate the "last" break in the series. A test for out of sample stability based on a real time monitoring rule is provided by Chu, Stinchcombe and White (1996). Recently Pesaran and Timmermann (1999) show that in the presence of structural breaks by taking into account observations across different regimes we reduce the forecast error variance, but we increase the bias and suggest how to find an optimal observational bandwidth.

While there is a very extensive literature on testing the null of zero versus one break and on the estimation of a single break, the literature on the testing and estimating multiple breaks is much more limited. Sequential tests for the null of  $m$  versus  $m + 1$  breaks are provided in Bai and Perron (1998) and in Bai (1997), the null of zero versus one break is sequentially tested over the different subsamples. Improving upon this, Bai (1999) proposes a sequential likelihood ratio test for the null of  $m$  versus  $m + 1$  breaks constructed using all the sample observations. Needless to say, sequential tests that avoid splitting the sample should have better finite sample power properties.

The purpose of these sequential testing procedures is to obtain a consistent estimate of all

---

<sup>1</sup>Since then, several papers have considered test for unit roots that allow for the presence of broken trend (see e.g. Banerjee, Lumsdaine and Stock (1992), Andrews and Zivot (1992), Lumsdaine and Papell (1997)). More recently, Leybourne, Mills and Newbold (1998) found that a break occurring "early" in the series may cause spurious rejection of the null of unit root.

the break fractions as well as to obtain a consistent estimate of the true number of breaks. If at any step of the procedure we can ensure that the probability of failing to reject the null when is false approaches zero, we can also ensure that the probability of underestimating the number of breaks is asymptotically zero. On the other hand, in order to ensure that the probability of overestimating the number of breaks is asymptotically zero, we need to ensure that at any stage the probability of rejecting the null when is true approaches zero. Bai (1999) provides an analytical expression for the critical values of his test and shows that by letting the critical values grow with the sample size, at a rate slower or equal to  $T$ , we can have a procedure that, at any stage, has zero asymptotic size and unit asymptotic power. However, in practice there are infinitely many sequences of critical values ensuring size approaching zero and power approaching one, so that it is not easy to decide which sequence of critical values should be used.

The object of this paper is to propose a simple sequential almost sure rule ensuring that, in large samples, both the probability of overestimating and underestimating the number of breaks is zero. As an immediate consequence, this rule provides a strong consistent estimate of the true number breaks. Our simple procedure is based on a straightforward application of the (functional) law of the iterated logarithm, (F)LIL.

We shall analyze the case of shifts in the mean of a possibly nonlinear process, allowing for dependent and heterogeneous observations. We shall begin by considering the following statistic:  $V_{ST} = \sup_{r \in [\gamma, 1-\gamma]} d_T^{-1} \hat{\sigma}^{-1} |\sum_{t=1}^{[Tr]} (X_t - \bar{X})|$ , with  $d_T = T^{1/2} \sqrt{2 \log \log T}$ ,  $\bar{X}$  is the sample mean, and  $\hat{\sigma}^2$  is a strong consistent estimator for the long-run variance matrix. Under the null of no breaks, the limsup of  $V_{ST}$  is almost surely less than 1/2. Under the alternative of at least one break, the statistic diverges at rate  $\frac{T^{1/2}}{\sqrt{l_T \log \log T}}$ , where  $l_T$  is the lag truncation parameter used in the estimation of the long-run covariance. If we get a number below the bound, we decide in favor of zero breaks and stop. Otherwise we proceed to estimate the first break, using all observations. We strengthen Bai's weak consistency result into a strong consistency result. More precisely we shall show that the estimated break fraction is  $T^{1-\eta}$ -strongly consistent for one of the true break fractions, with  $\eta > 0$ , arbitrarily small. Such a rate is slower than the  $T$ -rate obtained by Bai (1997) in the weak consistency case, but it suffices for our purposes. We then recompute the statistic above, by using estimators of the sample means over the proper

subsamples. In this way, at any step of the procedure we use all the observations in the sample. If we get a value below the bound we stop. otherwise we proceed to estimate a second break, by using all observations but those close to the break we previously estimated, thus ensuring we won't estimate the same break twice. This second break is also  $T^{1-\eta}$ -consistent for some true break, different from the one previously estimated. Then we shall proceed until we fail to reject the null of say  $m$  breaks versus  $m+1$  breaks. In this case we set the estimated number of breaks,  $m_T$ , equal to  $m$ , in fact at any step the rejection or not of the null depends on the sample size  $T$ . As at any step, in large samples, the probability of accepting (rejecting) the null when is correct (wrong) is one,  $m_T$  is a strongly consistent estimator of the true number of breaks.

Furthermore an estimator for the long-run variance which is consistent also in the presence of neglected breaks is proposed. The consistency under the alternative is attained by using an estimator of the mean based on an appropriate window of observations. This ensures that the test statistic diverges at rate  $\frac{\sqrt{T}}{\sqrt{2 \log \log T}}$  (instead of  $\frac{T^{1/2}}{\sqrt{l_T \log \log T}}$ ), under the alternative.

As the consistent estimation of the number of breaks depends on the fact that the size approaches zero at a proper rate, consistent estimation can be obtained via the Schwartz criterion (or BIC). In fact, broadly speaking, the Schwartz criterion amounts to comparing the likelihood-ratio test statistic with critical values diverging to infinity, where the speed of divergence is determined by the penalty term. Yao (1988) provided consistent estimation of the number of breaks, via the Schwartz criterion, under the assumption of normal innovations. More recently Liu, Wu, and Zidek (1997, hereafter LWZ) propose the use of a modified Schwartz criterion, characterized by a more severe penalty term, and so extend Yao's result to the case of identically independent observations, plus an additional condition that they call local exponential boundedness (LWZ, 1997, p.511). Approaches based on the Schwartz criterion require the estimation of all breaks at once; in fact for any given number of subsamples we obtain the breaks estimator by least squares and then we choose the number of subsamples (that plays the role of dimension of the model), by minimizing the (modified) Schwartz criterion. The potential advantages of our sequential approach over the Yao's and LWZ's approach are (i) we can allow for dependence and heterogeneity, (ii) we can estimate one break at time, which is computational less demanding, (iii) we provide a strongly, instead of weakly, consistent estimator of the true

number of breaks.

The sequential procedure has been applied to the same Eurodollar interest rate data set investigated by Ait-Sahalia (1996a-b). Multiple breaks have been detected over the period 1973-1995, in line with *ex post* revisitation of the US monetary policy. While the autocorrelation computed on the raw data displays a long memory type behavior, the one computed taking into account the shifts in the mean displays a short memory type behavior.

This paper is organized as follows. In Section 2 we describe our set-up, stating the assumptions that will be used in the sequel, and describing the testing procedure. Section 3 presents the asymptotic behavior of the procedure. Section 4 reports some Monte Carlo experiments. Section 5 applies the procedure to the weekly eurodollar rate. Finally, some concluding remarks are given in Section 6. All proofs are collected in the Appendix.

## 2 Set-up

### 2.1 Model and Assumptions

We shall focus on data generating processes (DGPs) characterized by shifts in the mean, such as

$$X_t = \mu_0 1\{k_0 + 1 \leq t \leq k_1\} + \mu_1 1\{k_1 + 1 \leq t \leq k_2\} + \dots \quad (1)$$

$$\dots + \mu_{m_0} 1\{k_{m_0} + 1 \leq t \leq k_{m_0+1}\} + \epsilon_t$$

where  $1\{\cdot\}$  denotes the indicator function, equal to 1 if the argument is true, and zero otherwise, hereafter  $k_0 = 0, k_{m_0+1} = T, \forall m_0 \geq 0$ .

In the sequel we shall need the following assumptions:

A1:  $\{\epsilon_t\}_{t=-\infty}^{\infty}$  is a zero mean, real-valued process, such that  $\sup_t E(|\epsilon_t|^{2(2+\delta)}) \leq C_1 < \infty, \delta > 0$ .

A2:  $\{\epsilon_t\}_{t=-\infty}^{\infty}$  is  $\alpha$ -mixing with size  $-\frac{4(4+\delta)}{\delta}$ , where  $\delta$  is the same positive constant defined in

A1. Also  $\lim_{T \rightarrow \infty} \frac{\sum_{t=1}^T E(\epsilon_t^2)}{T} = \sigma^2$ .

A3: There exists a constant  $0 < \sigma_0^2 < \infty$ , such that

$$|T^{-1} E\left(\left(\sum_{k=l+1}^{l+T} \epsilon_k\right)^2\right) - \sigma_0^2| \leq C_2 T^{-\psi}$$

where  $\psi > 0$  and  $C_2$  is independent of  $l$ . Note that, in general,  $\sigma_0^2 \neq \sigma^2$ , where  $\sigma^2$  is defined as in A2.

A4:  $\forall i = 0, \dots, m_0, k_{i+1} - k_i > \gamma T$ , for some  $\gamma > 0$ , and  $k_0 = 0, k_{m_0+1} = T$ .

A5:  $|\mu_i - \mu_{i+1}| \geq \Delta > 0, \forall i$ .

Assumptions A1 and A2 are standard memory and moment conditions; as usual there is a trade-off between the degree of heterogeneity and the degree of dependence we can allow for. A3 is a condition on the rate of growth of the variance of the partial sums of  $\epsilon_t$ , that is trivially satisfied in the covariance stationary case (see Corollary 2.2, Corradi 1999). Note that we do not constrain the innovation process to be linear, so allowing for a variety of nonlinear models. Assumptions A4 and A5 put some restrictions on the type of breaks we can allow for. A4 rules out the possibility that breaks are too close each other, that is we impose that the number of observations between two successive break points grows at rate  $T$ . Finally A5 requires that the magnitude of the breaks does not shrink as  $T$  gets large.

## 2.2 Testing Procedure

We now describe the proposed statistics and the steps of the sequential rule for detecting the number of breaks. The procedure starts by testing the null of the absence of breaks versus the presence of at least a break in the mean:

$$H_0^1 : \text{no breaks}$$

versus

$$H_A^1 : \text{at least one break.}$$

Thus under  $H_0^1$ , the DGP is as in (1), with  $m_0 = 0$ , while under  $H_A^1$  the DGP is as in (1), but with  $m_0 > 0$ .<sup>2</sup> In order to test  $H_0^1$  against  $H_A^1$ , we shall rely on the following statistic:

$$V_{ST}^1 = \sup_{r \in [\gamma, 1-\gamma]} \frac{1}{\hat{\sigma}_T} d_T^{-1} \left| \sum_{t=1}^{\lfloor Tr \rfloor} (X_t - \bar{X}) \right| \quad (2)$$

---

<sup>2</sup>The upper-script in the test hypotheses and in the statistics indicates the step of the procedure or differently the minimum number of breaks under the alternative.

where  $\gamma$  is defined as in A4,  $\bar{X} = T^{-1} \sum_{t=1}^T X_t$ ,  $d_T = T^{1/2} \sqrt{2 \log \log T}$ , and  $\hat{\sigma}_T^2$  is an HAC estimates of the variance, defined as:

$$\hat{\sigma}_T^2 = \frac{1}{T} \sum_{t=1}^T (X_t - \bar{X})^2 + \frac{2}{T} \sum_{s=1}^{l_T} w_s \sum_{t=s+1}^T (X_t - \bar{X})(X_{t-s} - \bar{X}) \quad (3)$$

with  $w_s$  the usual Bartlett window, i.e.,  $1 - \frac{s}{l_T+1}$ .

The statistic in (2) is the supremum of the partial sum of the residuals with respect to all the possible break points, taking into account that because of A4, the first break fraction cannot be smaller than  $\gamma$  and the last break fraction cannot be larger than  $1 - \gamma$ . The decision rule of the sequential procedure will be to stop if we get a value below or equal to  $\frac{1}{2}$  and then to decide in favor of no breaks. Otherwise we shall proceed to estimate the first break. As we shall show in proposition 3.1 below, in large sample the probability of choosing  $H_0^1$  when correct (resp. wrong) is one (resp. zero).

Having assessed the presence of at least one shift in the mean of the process, the first break is estimated by finding the sample location of the break which minimizes the sum of the residuals as in Bai (1997), that is:

$$\begin{aligned} \hat{k}^1 &= \underset{k=\lceil \gamma T \rceil, \dots, \lfloor (1-\gamma)T \rfloor}{\operatorname{argmin}} S_T(k) \\ &= \underset{k=\lceil \gamma T \rceil, \dots, \lfloor (1-\gamma)T \rfloor}{\operatorname{argmin}} \left\{ \sum_{t=1}^k (X_t - \mu(k/T))^2 + \sum_{t=k+1}^T (X_t - \mu^*(k/T))^2 \right\} \end{aligned} \quad (4)$$

where  $\mu(k/T) = \frac{1}{k} \sum_{t=1}^k X_t$  and  $\mu^*(k/T) = \frac{1}{T-k} \sum_{t=k+1}^T X_t$ . In Theorem 3.4 we will show that the estimated break is a  $T^{1-\eta}$  strong consistent estimator of the largest break in the sample, where with "largest" we mean that asymptotically it is responsible for the largest reduction in the sum of squared residuals.

Once we have identified the location of the first break, we test for the presence of additional breaks, conditional to the one already identified, i.e.:

$$H_0^2 : \text{only one break}$$

versus

$$H_A^2 : \text{at least two breaks}$$

by using the following statistic:

$$V_{ST}^2 = \frac{1}{\hat{\sigma}_T^2} d_T^{-1} \sup_{r \in [\gamma, 1-\gamma]} \left| \sum_{t=1}^{[Tr]} (X_t - \hat{X}_1 \mathbf{1}\{1 \leq t \leq \hat{k}^1\} - \hat{X}_2 \mathbf{1}\{\hat{k}^1 + 1 \leq t \leq [Tr]\}) \right|$$

where  $\hat{X}_1, \hat{X}_2$  are the means in the two subsamples which have been computed with respect to estimated break points, and  $\hat{\sigma}_T^2$  is as in equation (3), but for the fact that we replace  $X_t - \bar{X}$  with  $X_t - \hat{X}_1$  for  $t = 1, \dots, \hat{k}^1$  and with  $X_t - \hat{X}_2$  for  $t = \hat{k}^1 + 1, \dots, T$ .

As above, if we get a value for the statistic  $V_{ST}^2$  below or equal to  $\frac{1}{2}$  we stop and decide in favor of only one break.

If instead the statistic results larger than  $\frac{1}{2}$ , then we move on to the estimation of a second shift in the mean, taking into account the one already identified in the first step. We locate it by minimizing the sum of squared residuals with respect to the location of the second break point; in this estimation step we use all the observations in the sample but for a proper interval around the first break, so we do not rely on the idea of splitting the sample into different sub-samples as suggested by Bai (1997).  $A4$  and the  $T^{1-\eta}$ -consistency of  $\hat{k}^1/T$  for one of the true break fraction, ensure that we neither re-estimate the same break, nor we neglect some of the breaks. We shall show in the next session that  $\hat{k}^2/T$  is almost surely  $T^{1-\eta}$ -consistent too, for some break fraction, different from  $\tau_0^1$ .

Again, the statistics  $V_{ST}^3$  is computed by using the residuals of  $X_t$  from the relevant sample means over the different subsamples; we decide in favor of only two breaks if we get a value for  $V_{ST}^3$  below or equal to  $\frac{1}{2}$ . The procedure finally stops, when we accept the null of  $m$  breaks versus the alternative of more than  $m$  breaks.

Critical for the small sample performance of the decision rule is the use of an appropriate estimator of the long run variance. Under the null, the estimator is strongly consistent for the true variance, but it is no longer consistent under the alternative. In presence of neglected breaks, any HAC estimator of the covariance matrix diverges at rate  $l_T$ , where  $l_T$  is the lag truncation parameter, both in presence of dependent or martingale difference innovations. The inconsistency of the variance estimator stems from the fact that, in the presence of neglected breaks, the sample mean, constructed using all the observations, is no longer consistent for



$E(X_t)$ ,  $t = 1, 2, \dots, T$ . In the application section, we discuss how the effect of neglected breaks on the autocoraviance function can be confused with a long memory behavior.

The upward bias of the variance estimator reduces the power of the test, both in finite and large samples. To overcome this problem, we propose a new estimator of the variance, based on the deviations of each realizations with respect to a local mean, which is computed by averaging the observations in a given neighborhood. The use of a local mean, under proper conditions on the size of the neighborhood, allows to obtain a consistent estimate of the variance also under the alternative.

### *Spurious Breaks*

Detection of spurious breaks in the presence of integrated error is a known problem, as pointed out by Nunes, Kuan and Newbold (1995, 1996). Vogelsang (1998) proposes statistics that do not detect spurious breaks in the case of integrated error. However he deals with the case of only one break, and his purpose is not a consistent estimator of the number of breaks. We now suggest a simple device for distinguishing between a rejection of the null due to the existence of breaks and a rejection due to integrated errors. Suppose we have rejected the null of zero breaks versus the alternative of at least one break. The divergence of the statistics can be instead due to the fact that  $X_t$  is an integrated process.<sup>3</sup> So once we have rejected the null, we want to able to know whether the rejection was indeed due to the presence or breaks or was instead due to integratedness. We can proceed in the following way. Construct the auxiliary statistic:  $Z_T = \frac{1}{T^2} \sum_{t=1}^T (X_t - \bar{X})^2$ . If  $X_t$  is  $I(1)$ , then  $Z_T = \frac{1}{T^2} \sum_{t=1}^T (X_t - \bar{X})^2 = \frac{1}{T} \sum_{t=1}^T (\frac{1}{\sqrt{T}} \sum_{j=1}^t \epsilon_j)^2 \Rightarrow \sigma^2 \int_0^1 W_\mu^2(s) ds$ , where  $W_\mu$  is a demeaned standard Brownian motion. We can use as an estimator of  $\sigma^2$ ,  $\hat{\sigma}_T^2 = \frac{1}{T} \sum_{t=1}^T (\Delta X_t)^2$  in order to construct the percentile of the limiting distribution above.<sup>4</sup> If  $X_t$  is  $I(0)$  with breaks, then  $TZ_T$  converges in probability to a constant and so  $Z_T$  vanishes to zero at rate  $T^{-1}$ . Thus if we get a value for the statistic

---

<sup>3</sup>In fact a statistic similar  $V_{ST}^1$  has been used in Corradi (1999) to distinguish between  $I(0)$  and  $I(1)$  processes with no breaks. Under the  $I(1)$  alternative the statistic diverges.

<sup>4</sup> $\hat{\sigma}_T^2$  in the dependent case is not consistent for the true long run variance. However we avoid to use an HAC estimator as, if  $X_t$  is  $I(0)$  with breaks and  $\epsilon_t$  follows an ARMA process, then  $\hat{\sigma}_T^2$  would approach zero in probability as  $T \rightarrow \infty$ .

below say the 5% – percentile, we decide in favor of  $I(0)$  with breaks, otherwise we decide in favor of  $I(1)$  and stop searching.

### 3 Asymptotic behavior

Having described the procedure, we now show that the sequential rules described in the previous sections indeed provide an almost surely consistent estimator of the true number of breaks.

#### *Testing for the absence of breaks*

The asymptotic behavior of the statistic  $V_{ST}^1$ , defined in (2), is summarized in the following result:

#### **Proposition 3.1**

Let A1-A5 hold, let  $l_T \rightarrow \infty$  as  $T \rightarrow \infty$ , and  $l_T = O(T^{1/4}(\log T)^{-1/(4-\lambda)})$ , for some  $\lambda > 0$ . Under  $H_0^1$ , almost surely,

$$(i) \limsup_{T \rightarrow \infty} V_{ST}^1 = \frac{1}{2}$$

Under  $H_A^1$ ,

$$(ii) P \left( \lim_{T \rightarrow \infty} \frac{T^{1/2}}{(l_T \log \log T)^{1/2}} V_{ST}^1 > 0 \right) = 1.$$

From the proposition above, we see that to decide in favor of  $H_0^1$ , no breaks, when we get a value for  $V_{ST}^1$  below or equal to  $\frac{1}{2}$  and otherwise to decide in favor of  $H_A^1$ , at least one break, provides a completely consistent, almost sure rule. We also see that, under the alternative, the statistic diverges at rate  $\sqrt{\frac{T}{l_T \log \log T}}$ , thus, in finite sample, a "small" lag truncation parameter will improve the power, although it may deteriorate the size, whenever the innovations show a high degree of dependence.

As we pointed out in the previous section, the fact that the variance estimator is not consistent for the true variance under the alternative is likely to be one of the main reasons for the low power of tests for structural breaks. The inconsistency under the alternative occurs both in the case of martingale difference innovations and in that of dependent innovations, as

the denominator diverges at rate  $l_T$ , the size of the Bartlett window, and so the statistic diverges at rate  $\sqrt{T}/\sqrt{l_T \log \log T}$ , instead that at rate  $\sqrt{T}/\sqrt{\log \log T}$ . As shown in the proof, this is due to the fact that under the alternative  $\bar{X}$  is not consistent for  $E(X_t), \forall t$ . We then provide a modified version of the statistic  $V_{ST}^1$ , based on a "local" estimator of the sample mean, which ensures the consistency of the variance estimator under both the null and the alternative. Let denote  $V_{SL,T}^1$  the modified version, where:

$$V_{SL,T}^1 = \sup_{r \in [\gamma, 1-\gamma]} \frac{1}{\hat{\sigma}_{L,T}} d_T^{-1} \left| \sum_{t=1}^{\lfloor Tr \rfloor} (X_t - \bar{X}) \right| \quad (5)$$

with

$$\hat{\sigma}_{L,T}^2 = \frac{1}{T} \sum_{t=1}^T (X_t - \bar{X}^t)^2 + \frac{2}{T} \sum_{s=1}^{l_T} w_s \sum_{t=s+1}^T (X_t - \bar{X}^t)(X_{t-s} - \bar{X}^{t-s}) \quad (6)$$

with  $w_s = 1 - \frac{s}{l_T+1}$  and

$$\bar{X}^t = \begin{cases} \frac{1}{T^\xi} \sum_{j=1}^{T^\xi} X_j & \text{if } 1 \leq t \leq T^\xi \\ \frac{1}{2T^\xi} \sum_{j=t-T^\xi}^{t+T^\xi} X_j & \text{if } T^\xi + 1 \leq t \leq T - T^\xi \\ \frac{1}{T^\xi} \sum_{j=T+1-T^\xi}^T X_j & \text{if } T - T^\xi + 1 \leq t \leq T \end{cases} \quad (7)$$

and  $\bar{X}^{t-s}$  is defined in an analogous way. Thus, as an estimator of the mean of  $X_t$  we use an average over  $T^\xi$  observations around  $t$ . We shall show that  $\hat{\sigma}_{L,T}^2$  is consistent for  $\sigma_0^2$  under both the null and the alternative, provided we properly choose  $\xi$ . We have:

### Proposition 3.2

Let A1-A5 hold, let  $l_T \rightarrow \infty$ , as  $T \rightarrow \infty$  and  $l_T = O(T^{1/4}(\log T)^{-1/(4-\lambda)})$ , for some  $\lambda > 0$ ,  $\frac{l_T T^\xi \log \log T}{T} \rightarrow 0$ , and  $\frac{T^\xi}{l_T^2 \log \log T^\xi} \rightarrow \infty$ .

Under  $H_0^1$ , almost surely

$$(i) \quad \limsup_{T \rightarrow \infty} V_{SLT}^1 = 1/2$$

Under  $H_A^1$ ,

$$(ii) \quad P \left( \lim_{T \rightarrow \infty} \frac{T^{1/2}}{(\log \log T)^{1/2}} V_{SLT}^1 > 0 \right) = 1.$$

Note that differently from the result in proposition 3.1, the length of the lag-window does not affect the rate at which the statistic goes to infinity under the alternative, as a result of using a

consistent estimator of the variance. Following the described decision rule, if the statistic results above the threshold level, we proceed to locate the first break  $\hat{k}^1$  as in (4).

Using the following identity

$$\mu^* \left( \frac{k}{T} \right) = \frac{T}{T-k} \bar{\mu} - \frac{k}{T-k} \mu \left( \frac{k}{T} \right),$$

where  $\bar{X}$  is the sample average over the whole interval and after some algebraic manipulations, the objective function in (4) can be expressed more properly as

$$\frac{S_T(k)}{T} = \frac{\sum_{t=1}^T X_t^2}{T} - \frac{k}{T-k} \mu \left( \frac{k}{T} \right)^2 - \frac{T}{T-k} \bar{\mu}^2 + 2 \frac{k}{T-k} \bar{\mu} \mu \left( \frac{k}{T} \right),$$

this formulation will be largely exploited in the sequel. We now show that  $\hat{k}^1/T = \hat{\tau}^1$  is almost surely  $T^{1-\eta}$ -consistent for  $\tau_0^1$ , where  $\tau_0^1$  is one of the true break fraction. The following result, ensuring the uniform convergence of the objective function, is necessary.

### Lemma 3.3

Under A1 – A2 and A4 and for the estimation of the first break,

$$\sup_{\tau \in [\gamma, 1-\gamma]} \left| \frac{1}{T} S_T \left( \frac{[T\tau]}{T} \right) - U(\tau) \right| \xrightarrow{a.s.} 0$$

where

$$U(\tau) = \sigma^2 + \sum_{i=1}^{m_0+1} \mu_i^2 (\tau_i - \tau_{i-1}) - \frac{\tau}{1-\tau} \mu(\tau)^2 - \frac{1}{1-\tau} \mu^2 + 2 \frac{\tau}{1-\tau} \mu \mu(\tau) \quad (8)$$

with

$$\mu = \sum_{i=1}^{m_0+1} \mu_i (\tau_i - \tau_{i-1})$$

and

$$\mu(\tau) = \frac{1}{\tau} \left[ \sum_{i=1}^{i^*} \mu_i (\tau_i - \tau_{i-1}) + \mu_{i^*+1} (\tau - \tau_{i^*}) \right],$$

with  $\tau_{i^*}$  denoting the last break point before  $\tau$ .

Before stating the consistency result, we need an additional assumption:

A6:  $\exists \{\tau_0^1, \dots, \tau_0^m\}$  with  $\tau_0^i = \tau_j$ , with  $i$  not necessarily equal to  $j$  such that  $U(\tau_0^1) < U(\tau_0^2) \dots < U(\tau_0^m)$ , and for  $0 < n < m_0$ ,  $U^n(\tau_0^{n+1}) < U^n(\tau_0^{n+2}) \dots < U^n(\tau_0^{m_0})$ .<sup>5</sup>

---

<sup>5</sup>The first break we estimate is not an estimator of first break, but rather of the break that most contributes to the reduction of the sum of squared residuals.

This assumption is rather standard and essentially requires that there is a ranking among the break points in terms of their contribution to the minimization of the objective function. Broadly speaking it is a identification assumption which ensures that we first detect the break point that most contributes to the reduction of the sum of the square residuals.

**Theorem 3.4**

Let A1 – A2 and A4 – A6 hold, then

$$\hat{\tau}^1 - \tau_0^1 = O_{a.s.}(T^{-(1-\eta)})$$

for  $\eta > 0$  arbitrarily small,  $\hat{\tau}^1 = \hat{k}^1/T$ , with  $\hat{k}^1$  defined as (4).

From the theorem above we see that  $\hat{k}^1/T$  converges almost surely, at rate  $T^{1-\eta}$  to  $\tau_0^1$ , the break that most contributes to the reduction of the sum of square errors, or differently the break dominating the others in terms of relative magnitude of the shift. Bai (proposition 2, 1997) shows weak consistency at rate  $T$ . Thus the price that we pay in order to get strong consistency, is a slightly slower rate of convergence,  $T^{1-\eta}$  instead of  $T$ , with  $\eta > 0$  arbitrarily small. In fact, for our purposes, we just need convergence at a rate faster than  $\sqrt{T \log \log T}$ . As it is shown more in detail in the proof, the reason why we get  $T^{1-\eta}$ , rather than  $T$  consistency, is that, in order to ensure almost sure convergence, we need to let  $M$ , which is the minimum distance from the true break point after which the values of the objective function are almost surely increasing, to depend on  $T$ , that is we have to let  $M = M_T = O(T^\eta)$ , while  $M$  is a constant in proposition 2 of Bai (1997).

*Testing for the presence of only a break*

Given  $\hat{k}_1$  we construct  $V_{S,T}^2$ , and test the null of one break versus the alternative of at least two breaks. We have that the following proposition holds:

**Proposition 3.5**

Let A1-A6 hold, let  $l_T \rightarrow \infty$ , as  $T \rightarrow \infty$ , and  $l_T = O(T^{1/4}(\log T)^{-1/(4-\lambda)})$ , for some  $\lambda > 0$ .

Under  $H_0^2$ , almost surely

$$(i) \limsup_{T \rightarrow \infty} V_{S,T}^2 \leq 1/2$$

Under  $H_A^2$ ,

$$(ii) P \left( \lim_{T \rightarrow \infty} \frac{T^{1/2}}{(l_T \log \log T)^{1/2}} V_{S,T}^2 > 0 \right) = 1.$$

Thus at the second step, to decide in favor of  $H_0^2$  if we get values below or equal to  $\frac{1}{2}$ , and decide in favor of  $H_A^2$  otherwise, provides an almost sure rule. From Part (ii) of the proposition we note that under the alternative the statistic diverges at rate  $\sqrt{T}/\sqrt{l_T \log \log T}$ . If we would have scaled the statistic using  $\hat{\sigma}_{L,T}^2$  as defined in (6), properly modified to take into account the already estimated break point and with  $T^\xi/l_T^2 \log \log T \rightarrow \infty$ ,  $l_T T^\xi \log \log T/T \rightarrow 0$  as  $T \rightarrow \infty$ , then under the alternative we should have had divergence at rate  $\sqrt{T}/\sqrt{\log \log T}$ , by the same argument used in the proof of proposition 3.2. As before, if the statistic is above the bound, the estimation of an additional shift in the mean has to be considered.

Here we look at the general case of estimating the  $n + 1 - th$  break,  $0 < n < m_0$ , with  $m_0$  being the true number of breaks. We need to take into account the  $n$  breaks already estimated, that is  $\{\hat{\tau}^1, \dots, \hat{\tau}^n\}$ . The objective function,  $S_T^n(k)$ , for the estimation of the  $n + 1 - th$  break, conditional to the previous  $n$ , is given for a generic  $k$ , with  $k \in (\hat{k}^i, \hat{k}^{i+1})$   $i \leq m_0$ , is given by:<sup>6</sup>

$$\begin{aligned} S_T^n(k) &= \sum_{t=1}^T x_t^2 - \left( \hat{k}^{i+1} - \hat{k}^i \right) \left[ \frac{k - \hat{k}^i}{\hat{k}^{i+1} - k} \mu(\hat{k}^i, k)^2 + \frac{\hat{k}^{i+1} - \hat{k}^i}{\hat{k}^{i+1} - k} \mu(\hat{k}^i, \hat{k}^{i+1})^2 \right. \\ &\quad \left. - 2 \frac{k - \hat{k}^i}{\hat{k}^{i+1} - k} \mu(\hat{k}^i, k) \mu(\hat{k}^i, \hat{k}^{i+1}) \right] \\ &\quad - \sum_{j=1}^{i-1} \left( \hat{k}^{i+1} - \hat{k}^i \right) \mu(\hat{k}^j, \hat{k}^{j+1})^2 - \sum_{j=i+1}^n \left( \hat{k}^{i+1} - \hat{k}^i \right) \mu(\hat{k}^j, \hat{k}^{j+1})^2 \end{aligned} \quad (9)$$

where  $\mu(k, k^*) = \frac{1}{k^* - k} \sum_{t=k+1}^{k^*} X_t$ . Again the uniform convergence of the objective function with respect to the estimated parameters is crucial:

### Lemma 3.6

Under  $A1 - A2$  and  $A4$  and for  $n = 1, \dots, m_0$

$$\sup_{\tau \in [\gamma, 1-\gamma]} \left| \frac{S_T^n([T\tau])}{T} - U^n(\tau) \right| \xrightarrow{a.s} 0$$

---

<sup>6</sup>Remember that  $k^0 = 0$  and  $k^{m_0+1} = T$ .

where for  $\tau \in (\tau_i, \tau_{i+1})$ , we have that

$$\begin{aligned}
U^n(\tau) &= \sigma^2 + \sum_{i=1}^{m+1} \mu_i^2 (\tau_i - \tau_{i-1}) - (\tau_{i+1} - \tau_i) \left[ \frac{\tau - \tau_i}{\tau_{i+1} - \tau} \mu(\tau_i, \tau)^2 \right. \\
&\quad \left. + \frac{\tau_{i+1} - \tau_i}{\tau_{i+1} - \tau} \mu(\tau_i, \tau_{i+1})^2 - 2 \frac{\tau - \tau_i}{\tau_{i+1} - \tau} \mu(\tau_i, \tau) \mu(\tau_i, \tau_{i+1}) \right] \\
&\quad - \sum_{j=1}^{i-1} (\tau_{i+1} - \tau_i) \mu(\tau_j, \tau_{j+1})^2 - \sum_{j=i+1}^n (\tau_{i+1} - \tau_i) \mu(\tau_j, \tau_{j+1})^2
\end{aligned} \tag{10}$$

with  $\mu(\tau_j, \tau_{j+1}) = \lim_{T \rightarrow \infty} \frac{1}{k_{j+1} - k_j} \sum_{t=k_j+1}^{k_{j+1}} X_t$  with  $k_j = \lceil \tau_j T \rceil$  and similarly for  $\tau_{j+1}$ .

In the case of the estimator of the second break,  $n = 1$  let  $\hat{k}^{n+1} = \hat{k}^2$ , be defined as

$$\hat{k}^2 = \underset{k=\lceil \gamma T \rceil, \dots, \hat{k}^1 - \lceil \gamma T \rceil - 1, \hat{k}^1 + \lceil \gamma T \rceil + 1, \dots, \lceil (1-\gamma)T \rceil}{\operatorname{argmin}} S_T^1(k) \tag{11}$$

where  $\gamma$  is assumed in A4. Let  $\hat{\tau}_0^2 = \hat{k}^2/T$ , we have that the following corollary holds:

**Corollary 3.7**

Under A1 – A2 and A4 – A6

$$\hat{\tau}^2 - \tau_0^2 = O_{as}(T^{-(1-\eta)})$$

for  $\eta > 0$ , arbitrarily small.

Thus the second break, estimated conditional to the first one, converges almost surely to the second more prominent break, at rate  $T^{1-\eta}$  which is the same rate of the previously estimated break; we then proceed sequentially, until we fail to reject the null of additional breaks. The results can be easily extended to the case of a generic number of breaks.

*The total number of breaks*

The estimated number of breaks  $m_T$  is given by the number of breaks under the null, when we finally accept it. The following theorem ensures that  $m_T$  is strongly consistent for the true number of breaks:

**Theorem 3.8**

Let A1-A6 hold, let  $l_T \rightarrow \infty$ , as  $T \rightarrow \infty$ , and  $l_T = O(T^{1/4}(\log T)^{-1/(4-\lambda)})$ , for some  $\lambda > 0$ .

Then

$$P\left(\lim_{T \rightarrow \infty} m_T > m_0\right) = P\left(\lim_{T \rightarrow \infty} m_T < m_0\right) = 0$$

where  $m_0$  is the true number of breaks.

From Theorem 3.8, we see that  $m_T$  is a strongly consistent estimator of the true number of breaks.

**Remark:**

For completeness, we now define an additional statistic based on the average of partial sums, rather than on the supremum of partial sums. Consider:

$$V_{MT}^1 = \frac{1}{\hat{\sigma}_T^2} b_T^{-1} \sum_{t=1}^T \left| \sum_{j=1}^t (X_j - \bar{X}) \right|$$

and

$$V_{MT}^2 = \frac{1}{\hat{\sigma}_T^2} b_T^{-1} \sum_{t=1}^T \left| \sum_{j=1}^t (X_j - \hat{X}_1 1\{1 \leq j \leq \hat{k}^1\} - \hat{X}_2 1\{\hat{k}^1 + 1 \leq j \leq T\}) \right|$$

with  $b_T = T^{3/2} \sqrt{2 \log \log T}$ . By noting that  $V_{MT}^1 \leq V_{ST}^1$  and  $V_{MT}^2 \leq V_{ST}^2$  we have that under the null of zero breaks and the null of one break respectively,  $\limsup_{T \rightarrow \infty} V_{MT}^1 \leq 1/2$  and  $\limsup_{T \rightarrow \infty} V_{MT}^2 \leq 1/2$ , almost surely. Such a result is a substantial improvement over Corradi (1999, Theorem 4.1(i)) where for the statistic  $V_{MT}^1$  was provided a looser upper bound, equal to  $\frac{1}{\sqrt{3}} + \frac{1}{2}$ . Overall we expect that in finite sample the average statistic has smaller type I error, but larger type II error.

## 4 Findings from a Monte Carlo simulation exercise

In this section we analyze the finite sample performance of the testing procedure via a Monte Carlo exercise. In all the experiments we have chosen the parameter  $\xi$ , governing the number of observations used in the construction of the sample local mean (7) and (6) via cross validation, although we are aware of the fact that the asymptotic validity of the cross validation procedure holds only in the *iid* case. First we consider the behavior of  $V_{MT}$  and  $V_{ST}$  when there are no breaks, then in the case of one break and finally in the case of multiple (two and three) breaks.



We consider two sample sizes:  $T = 200$  and  $T = 400$ . In all the tables below, each entry denotes the (percentage) number of times we have chosen the number of breaks stated at the top of the column.

#### 4.1 DGPs with no breaks

We have generated  $X_t$  according to the following DGP:  $X_t = \epsilon_t = \rho\epsilon_{t-1} + u_t$  where the innovation  $u_t$  is independent normal with variance  $1 - \rho^2$  to ensure the same unconditional variance of the process  $X_t$  across experiments. The parameter  $\gamma$ , which regulates the minimal distance between adjacent break points (see A5) has been set equal to 0.05.<sup>7</sup> Three values of  $\rho$  have been considered.<sup>8</sup>

$T = 200$	$V_{SL}$			$V_{ML}$		
	$n = 0$	$n = 1$	$n \geq 2$	$n = 0$	$n = 1$	$n \geq 2$
$\rho = 0.0$	0.84	0.15	0.01	0.96	0.03	0.01
$\rho = 0.5$	0.80	0.17	0.02	0.96	0.04	0.00
$\rho = 0.9$	0.72	0.25	0.03	0.95	0.04	0.01

$T = 400$	$V_{SL}$			$V_{ML}$		
	$n = 0$	$n = 1$	$n \geq 2$	$n = 0$	$n = 1$	$n \geq 2$
$\rho = 0.0$	0.86	0.13	0.01	0.99	0.01	0.00
$\rho = 0.5$	0.83	0.16	0.01	0.98	0.02	0.00
$\rho = 0.9$	0.75	0.22	0.02	0.98	0.02	0.0

As the data have been generated with no breaks, the  $V_{SL}$  has a type I error around 0.15 in the case of uncorrelated of the errors, the type I error rises to above 0.20 in presence of high correlation in the data; the dimension of the sample size slowly improves the results. The

<sup>7</sup>All the computation were carried out in double precision FORTRAN, using a random generator from the NAG library. All the reported simulation results are based on 2.000 Monte Carlo replications.

<sup>8</sup>The lag truncation in the heteroskedasticity and autocorrelation consistent estimation of the covariance matrices has been chosen according to the automatic bandwidth of Andrews (1991).

statistics based on the supremum can wrongly detect the presence of one break, however the statistics never points to more than one break. On the other hand,  $V_{ML}$  is around 0.02 for  $T = 400$ , and around 0.04 for  $T = 200$ . This is not surprising as at any stage of the procedure  $V_{ML} \leq V_{SL}$ .

## 4.2 DGPs with a single break

We now consider DGPs with a single mean shift, i.e.  $X_t = \mu_0 1\{1 \leq t \leq k_1\} + \mu_1 1\{k_1 + 1 \leq t \leq T\} + \epsilon_t$  and the  $\epsilon_t$  term is an  $AR(1)$  process, i.e.,  $\epsilon_t = 0.5\epsilon_{t-1} + u_t$  where the innovation  $u_t$  is independent normal with variance  $1 - 0.5^2$ . For convenience we will fix  $\mu_0 = 0$  and we will consider two different values of the mean in the second subperiod,  $\mu_1 = 0.5, 1$  and three different positions of the break point in the sample,  $\frac{k_1}{T} = .2, .35, .5$ .

$T = 200$	$V_{SL}$			$V_{ML}$		
$\mu_1 = 0.5$	$n = 0$	$n = 1$	$n \geq 2$	$n = 0$	$n = 1$	$n \geq 2$
$\frac{k_1}{T} = 0.2$	0.33	0.56	0.11	0.89	0.11	0.00
$\frac{k_1}{T} = 0.35$	0.27	0.63	0.08	0.85	0.15	0.00
$\frac{k_1}{T} = 0.5$	0.21	0.71	0.08	0.80	0.20	0.00

$T = 200$	$V_{SL}$			$V_{ML}$		
$\mu_1 = 1$	$n = 0$	$n = 1$	$n \geq 2$	$n = 0$	$n = 1$	$n \geq 2$
$\frac{k_1}{T} = 0.2$	0.02	0.81	0.17	0.51	0.49	0.00
$\frac{k_1}{T} = 0.35$	0.02	0.85	0.12	0.40	0.60	0.00
$\frac{k_1}{T} = 0.5$	0.01	0.89	0.10	0.23	0.77	0.00

$T = 400$	$V_{SL}$			$V_{ML}$		
$\mu_1 = 0.5$	$n = 0$	$n = 1$	$n \geq 2$	$n = 0$	$n = 1$	$n \geq 2$
$\frac{k_1}{T} = 0.2$	0.20	0.73	0.07	0.82	0.18	0.00
$\frac{k_1}{T} = 0.35$	0.15	0.78	0.06	0.77	0.23	0.00
$\frac{k_1}{T} = 0.5$	0.07	0.88	0.05	0.68	0.34	0.00

$T = 400$	$V_{SL}$			$V_{ML}$		
$\mu_1 = 1$	$n = 0$	$n = 1$	$n \geq 2$	$n = 0$	$n = 1$	$n \geq 2$
$\frac{k_1}{T} = 0.2$	0.00	0.88	0.12	0.24	0.76	0.00
$\frac{k_1}{T} = 0.35$	0.00	0.90	0.09	0.13	0.86	0.00
$\frac{k_1}{T} = 0.5$	0.01	0.95	0.05	0.03	0.96	0.01

In presence of a single break, the rule based on the supremum performs extremely well, even with samples of relative short length, as typical with quarterly macroeconomic time series ( $T = 200$ ). The presence of break of small size,  $\mu_1 = 0.5$  (half of the standard error of the innovations) or positioned at the extreme of the sample slightly reduce the ability of the statistic of detecting the break. On the other hand,  $V_{MT}$  performs well only with relatively large shift and break point not too close to the beginning of the sample.

### 4.3 DGPs with multiple breaks

We now look at the performance of the statistics in the presence of multiple breaks, and we will perform two type of exercises. In the first one, we consider a DGP with two break points, i.e.  $X_t = \mu_0 1\{1 \leq t \leq k_1\} + \mu_1 1\{k_1 + 1 \leq t \leq k_2\} + \mu_2 1\{k_2 + 1 \leq t \leq T\} + \epsilon_t$  and the  $\epsilon_t$  term is an  $AR(1)$  with autoregressive parameter equal to 0.5. We keep fix the break dates at  $\frac{k_1}{T} = 1/3$  and  $\frac{k_2}{T} = 2/3$  and vary the relative size of the shift, measured as  $\frac{\mu_2 - \mu_1}{\mu_1 - \mu_0}$ . We will fix  $\mu_1 - \mu_0 = 1$  and we will consider samples of 200 and 400 observations.

$T = 200$	$V_{SL}$				$V_{ML}$			
	$n = 0$	$n = 1$	$n = 2$	$n \geq 3$	$n = 0$	$n = 1$	$n = 2$	$n \geq 3$
$\frac{\mu_2 - \mu_1}{\mu_1 - \mu_0} = 0.5$	0.01	0.67	0.31	0.01	0.03	0.96	0.00	0.00
$\frac{\mu_2 - \mu_1}{\mu_1 - \mu_0} = 1.0$	0.01	0.29	0.68	0.02	0.01	0.99	0.00	0.00

$T = 400$	$V_{SL}$				$V_{ML}$			
	$n = 0$	$n = 1$	$n = 2$	$n \geq 3$	$n = 0$	$n = 1$	$n = 2$	$n \geq 3$
$\frac{\mu_2 - \mu_1}{\mu_1 - \mu_0} = 0.5$	0.00	0.48	0.51	0.01	0.00	0.99	0.01	0.00
$\frac{\mu_2 - \mu_1}{\mu_1 - \mu_0} = 1.0$	0.00	0.09	0.88	0.03	0.00	0.99	0.01	0.00

We see that  $V_{ML}$  in all cases is able to detect only one break, although it practically never fails to detect it. On the other hand in the case of large shift (i.e. shift equal to the standard deviation of the innovation)  $V_{ST}$  is able to detect both breaks about 68% of times (for  $T = 200$ ) and 88% of times (for  $T = 400$ ); the procedure chooses either one or two breaks.

We now consider the behavior of  $V_{ST}$  when in the  $DGP$ s there are three breaks<sup>9</sup> located at  $\frac{k_1}{T} = 1/4$ ,  $\frac{k_2}{T} = 2/4$ ,  $\frac{k_3}{T} = 3/4$ . Three different parametrizations of the shift in mean have been analyzed.

$T = 200$	$V_{SL}$				
	$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n \geq 4$
$(\mu_0, \mu_1, \mu_2, \mu_3)$					
$(1, 2.5, 1.5, 0.5)$	0.00	0.06	0.48	0.45	0.00
$(1, 2, 3, 4)$	0.00	0.13	0.54	0.33	0.00
$(1, 2, 1, 2)$	0.09	0.29	0.13	0.46	0.03

<sup>9</sup>In the case of three breaks  $V_{MT}$  was able to detect only either one or two breaks, in all cases.

$T = 400$	$V_{SL}$				
$(\mu_0, \mu_1, \mu_2, \mu_3)$	$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n \geq 4$
(1,2.5,1.5,0.5)	0.00	0.00	0.20	0.79	0.01
(1,2,3,4)	0.00	0.02	0.25	0.73	0.00
(1,2,1,2)	0.03	0.15	0.06	0.74	0.02

In the first parameterization it is easier to detect all the breaks, as it is characterized by a large shift at the beginning of the sample; while the second and third parameterizations have all shifts of the same size. In particular the third scenario is the most difficult, given that difference between the subperiod means and the mean computed on the whole sample is equal to half standard deviation of the innovation.<sup>10</sup> A sample size of 200 observations seems too small for correctly detecting the number of breaks, as each subperiod has only fifty observations. However the procedure correctly detects the presence of multiple breaks. In the first two cases it tends to detect two breaks instead of three, while in the last case the procedure tends to find either a single break or three breaks. With the larger sample size, the rule is able to detect the correct number of breaks. As the errors display a relatively high degree of dependence ( $\rho = 0.5$ ), these results seem rather encouraging. Finally, as term of comparison, we repeat the experiment using an estimator of the variance based on a sample mean computed over the entire sample (i.e. we set  $\xi$  in (7) equal to one). We note that the resulting overestimation of the variance, induced by the neglected mean shift, slightly reduces the ability of the procedure of detecting multiple breaks.

<sup>10</sup>If we had only two breaks with  $\mu_0 = 1$ ,  $\mu_1 = 2$  and  $\mu_2 = 1$  then the moving estimates test of Chu, Hornik and Kuan (1995) would be probably the most appropriate test, given that it has been properly designed to detect a temporary parameters shifts.

$T = 200$	$V_S$				
$(\mu_0, \mu_1, \mu_2, \mu_3)$	$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n \geq 4$
(1, 2.5, 1.5, 0.5)	0.02	0.12	0.51	0.34	0.00
(1,2,3,4)	0.00	0.23	0.52	0.25	0.00
(1,2,1,2)	0.30	0.48	0.08	0.14	0.00

$T = 400$	$V_S$				
$(\mu_0, \mu_1, \mu_2, \mu_3)$	$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n \geq 4$
(1, 2.5, 1.5, 0.5)	0.00	0.03	0.23	0.74	0.00
(1,2,3,4)	0.00	0.06	0.29	0.65	0.00
(1,2,1,2)	0.15	0.44	0.05	0.36	0.00

From the findings reported above, we see that if there are no breaks  $V_{MT}$  never finds a break while  $V_{ST}$  tends to find a break in about 15% – 20% of the cases. On the other hand, if indeed there are structural breaks then  $V_{ST}$  is able to detect the correct number of breaks in most of the cases. Thus if  $V_{ST}$  detects zero break or more than one break we rely on such a findings. On the other hand if  $V_{ST}$  detects one break, we cannot draw a clear cut conclusion on whether there is a break or not. In such a case, it would be recommendable to test the null of zero versus one break using some other testing procedure, such as Bai (1997,1999) or Kuan, Nunes and Newbold (1995).

## 5 Are there mean shifts in the Eurodollar interest rate?

In finance theory, the pricing of a derivative security is based upon continuous time arbitrage arguments, stemming from the specification of a diffusion process for the underlying asset. To this end, the correct specification of the functional form of the drift and the volatility term of the underlying diffusion process is crucial for the correct pricing of the derivative security. Following this basic intuition, a large body of literature has paid attention to the issue of

testing for the correct specification of a diffusion process. Given that the drift and the variance uniquely determine the stationary density of the diffusion, a natural approach is to compare the parametric density, implied by a parametric specification of the drift and the variance, with a nonparametric estimator of the density. This is the approach proposed by Aït-Sahalia in a sequence of papers (1996a, 1996b) and applied to testing continuous time models of the spot interest rate.

Due to the usual change of measure argument the volatility term plays a crucial role in the pricing of derivatives, it is nevertheless true that the choice of the proper functional form for the volatility is not independent of the correct specification of the drift. Indeed one of the main findings of Ait-Sahalia (1996b), when applying his framework to the characterization of the properties of seven-day Eurodollar deposit rate, is that the linearity of the drift imposed in the literature is the main source of misspecification. To this end he proposed a non linear specification of the drift which is able to capture the mean reverting behavior in presence of large deviations, while being virtually zero around the mean of the process.

In light with those findings, the same data set of spot rate has been analyzed via the procedure described in the previous sections, in order to test for the possible presence of shifts in the mean. The data<sup>11</sup> set consists of seven day Eurodollar deposit rate, bid-ask midpoint, continuously compounded yield to maturity, from 01.06.1973 to 25.02.1995. The data has been sampled at weekly frequency, considering closing of the week value, and total sample size amounts of 1131 weekly observations. The weekly data are the solid line in Figure 1, while the autocorrelation function of the spot rate levels is displayed in Figure 2 with the solid line. The ACF exhibits an extremely slow decaying pattern, in fact the correlation seems to vanish at a linear rather than at a geometric one.

---

<sup>11</sup>The data are freely available from Aït-Sahalia's homepage, [www.princeton.edu/~yacine](http://www.princeton.edu/~yacine).

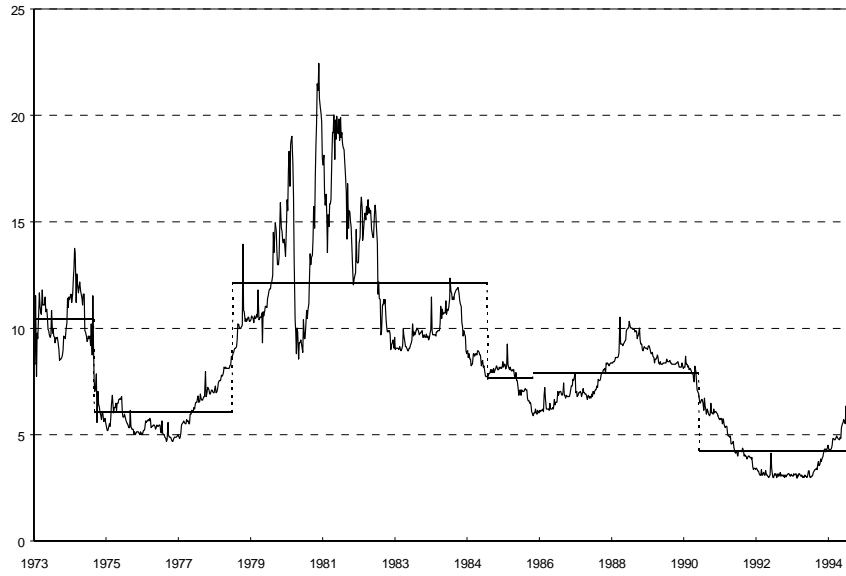


Figure 1 - Eurodollar spot rate

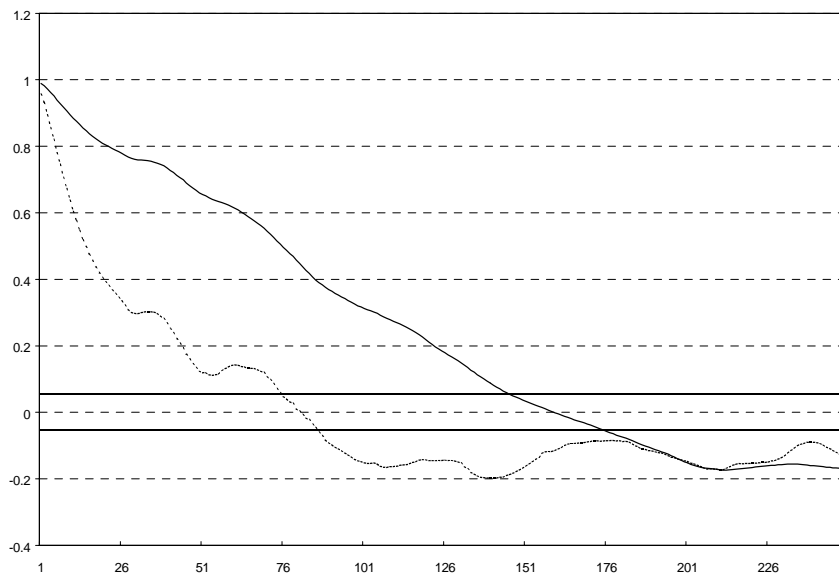


Figure 2 - Autocorrelation function (bounds equal to  $\pm 1.96T^{-1/2}$ ).

The sequential procedure for the detection of the number of breaks as well as for the locations of the break points, has been applied to the level interest rate, using the local mean



statistic  $V_{SL}$ , defined as in (5). The window length has been set equal to 320 periods ( $\xi = .41$ ) according to a cross-validation type of argument, while the length of the truncation lag in the computation of the unconditional variance has been set equal to 35, i.e.,  $l_T = 35$ . The results of the sequential procedure are reported in the following table.

$H_0$	$H_1$	$V_{SL}$
0	$\geq 1$	1.24
1	$\geq 2$	0.67
2	$\geq 3$	1.01
3	$\geq 4$	1.06
4	$\geq 5$	0.68
5	$\geq 6$	0.48

Thus we conclude in favor of five break points. In the table below, the first column displays the six different subperiods (in terms of the week of the year), while the second column displays the different means. The means over the different time spans are also plotted in Figure 1 with a dotted line.

Subperiods	Mean
1973/23 – 1975/1	10.414
1975/2 – 1978/36	6.057
1978/37 – 1985/24	12.119
1985/23 – 1986/35	7.644
1986/36 – 1991/4	7.883
1991/5 – 1994/7	4.238

The procedure detected the presence of five shifts in the mean of the spot Eurodollar rate from 1973 to 1995. Those dates are in line with *ex post* revisitation of the US monetary policy (e.g.: Romer and Romer, 1989; Rudebusch, 1995). The beginning of 1975 and middle of 1978 are two episodes discussed by Romer and Romer and the latter episodes is also included in the

so called "Romer dates." During the 1985 and 1986, there has been first a loosening of monetary policy, followed by a change of behavior in the end of 1986, as reported by Rudebusch (1995) in his study on the modification of the Federal Reserve interest rate target. Finally, the 1991 is middle point of a process of rate reduction started in the early 1989, with a FED target rate of 9.75%, and terminated at the end of 1992, with an interest rate level of 3%.

While the break dates are consistent with our economic priors, the validity of the analysis can be affected by the strong dependence detected in the data. To this end, the ACF has been recomputed after subtracting from the spot rate an estimator of the mean for each subsample; such ACF is plotted in Figure 2 with dashed line. Surprisingly the autocorrelations function shows a completely different pattern with respect to the one computed without considering the mean shifts. In particular, even if the correlation at lag one is still close to one, it seems to decay at a much quicker (exponential type) rate. The correlation displays some limited dynamic around one hundred lags, which, for weekly data, corresponds to a period of two years. So it clearly emerges that neglecting the mean shifts induces a spurious overestimation of the autocovariance and so a very slowly decay pattern of the autocorrelation function. Once we properly compute the sample mean, taking into account the breaks, the memory of the error of the process decays at an exponential-type rate, thus we can conclude that assumption A2, regulating the memory of the errors, is satisfied and our findings are reliable. Furthermore the results of the sequential procedure are not affected by the neglected breaks, given that the local mean has been utilized in the computation of the variance.

This empirical findings are somewhat related to recent literature which points out the problem of spurious identification of long memory due to neglected structural breaks (Diebold and Inoue, 1999).<sup>12</sup> In fact, as we already discussed, in the presence of neglected breaks the estimator of the variance of the partial sum is of the order  $O(Tl_T)$ , where  $l_T$  is the lag truncation parameter which is of order  $T^{1/4}(\log T)^{-1/(4-\lambda)}$ ,  $\lambda > 0$ . Thus the behavior of the series can be misleadingly confused with the behavior of a long memory process.

The correct specification of the drift term is crucial for the proper identification of the

---

<sup>12</sup>On the other hand Hidalgo and Robinson (1996) proposed a test for a single structural break in the presence of long memory errors.

volatility term characterizing the diffusion process; in this example, we show that the weekly eurodollar spot rate presents several mean shifts, which are in line with the change in the FED policy over the last twenty years. We have shown that neglecting those breaks can be an explanation for the strong dependence (of long memory type), which is usually observed in those data. In a recent paper Pritsker (1998) shows via monte carlo exercises, that in presence of strong dependence, Ait-Sahalia's specification procedure tends to overreject the null. Given the almost  $T$ -strong consistency of the break point, it may be appropriate to perform the Ait-Sahalia non parametric test over the different subsamples.

## 6 Conclusions

In this work, a simple sequential rule for detecting the number of shifts in the mean is proposed. We show that if, at any stage, we accept the null of  $l$  versus at least  $l + 1$  breaks whenever we get a value smaller than or equal to  $1/2$ , and we reject the null otherwise, then in large sample the probability of both type I and type II error is zero. The sequential procedure delivers a strongly consistent estimator of the number of shifts.

Critical for the performance of the sequential rule is the use of a consistent estimator for the long run variance, under both the null and the alternative. Under the alternative, standard HAC estimators of the variance deliver an over-estimation of the true variance of the process due to the neglected shifts in the mean. This can heavily jeopardize the ability of rejecting the null when it is false. We propose an estimator of the variance, based on deviations of the process from a properly defined local mean, which is a consistent estimator under both the null and the alternative, at any step of the procedure.

Some preliminary simulation results have shown that the procedure performs rather satisfactorily with samples of 400 observations, in the case of more than two breaks, and with samples of 200 observations in the case of at most two breaks. We have applied the procedure to the weekly Eurodollar deposit rate and we have detected five breaks, which are located in lines with changes in the U.S. monetary policy.

## 7 Appendix

In the proofs we will make extensive use to the following notation:  $1\{k, k_1\} = 1\{k \leq t \leq k_1\}$ ;  $d_T = \sqrt{T2 \log(\log(T))}$ .

PROOF OF PROPOSITION 3.1:

(i) The DGP under the null is:  $X_t = \mu_0 + \epsilon_t$ . We need to show that :

(a)

$$\limsup_{T \rightarrow \infty} \sup_{r \in [\gamma, 1-\gamma]} d_T^{-1} \left| \sum_{t=1}^{\lfloor Tr \rfloor} \epsilon_t - r \sum_{t=1}^T \epsilon_t \right| = \frac{1}{2} \sigma_0$$

where  $\sigma_0^2$  is as in A3, and  $\gamma$  is as in A4.

(b)  $\hat{\sigma}_T^2 \xrightarrow{a.s.} \sigma_0^2$ .

We begin by showing (a). We have

$$\begin{aligned} & \limsup_{T \rightarrow \infty} \sup_{r \in [\gamma, 1-\gamma]} d_T^{-1} \left| \sum_{t=1}^{\lfloor Tr \rfloor} \epsilon_t - r \sum_{t=1}^T \epsilon_t - (B(Tr) - rB(T)) \right| \leq \\ & \limsup_{T \rightarrow \infty} \sup_{r \in [\gamma, 1-\gamma]} d_T^{-1} \left| \sum_{t=1}^{\lfloor Tr \rfloor} \epsilon_t - B(Tr) \right| + \limsup_{T \rightarrow \infty} \sup_{r \in [\gamma, 1-\gamma]} r d_T^{-1} \left| \sum_{t=1}^T \epsilon_t - B(T) \right| = I + II \end{aligned}$$

where  $B$  is a Brownian motion with variance equal to  $\sigma_0^2$ . We show that  $I$  and  $II$  are  $o_{a.s.}(1)$ . Given A1-A3,  $I = o_{a.s.}(1)$  because of Theorem 1 in Eberlain (1986), once we have shown that (12) and (13) below hold:

$$\|E(S_T(m)|F_m)\|_2 \leq C \tag{12}$$

where  $S_T(m) = \sum_{k=m+1}^{m+T} \epsilon_k$ , and  $C$  is independent of  $m$ , and  $F_m = \sigma(\dots \epsilon_{m-1}, \epsilon_m)$  and uniformly in  $m$ ,

$$T^{-1} E|E(S_T(m)^2|F_m) - E(S_T(m)^2)| = O(T^{-\theta}) \tag{13}$$

for some  $\theta > 0$ .

We begin by showing (12). By Lemma 3.5 in McLeish (1975a) and by Minkowski's inequality,

$$\begin{aligned} \|E(S_T(m)|F_m)\|_2 & \leq \|E(\epsilon_{m+1}|F_m)\|_2 + \|E(\epsilon_{m+2}|F_m)\|_2 + \dots + \|E(\epsilon_{m+T}|F_m)\|_2 \\ & \leq 6\alpha_1^{1/2-1/2(2+\delta)} \|\epsilon_{m+1}\|_{2(2+\delta)} + \dots + 6\alpha_T^{1/2-1/2(2+\delta)} \|\epsilon_{m+T}\|_{2(2+\delta)} \end{aligned}$$

$$\leq 6 \sup_t \|\epsilon_t\|_{2(2+\delta)} \sum_{j=0}^{\infty} \alpha_j^{1/2-1/2(2+\delta)} \leq 6C_1^{1/2(2+\delta)} A \leq C$$

because of the bound in A1 and because, given A2,  $\sum_{j=1}^{\infty} \alpha_j^{1/2-1/2(2+\delta)} = A < \infty$ .

As for (13) it holds by the same argument used in the proof of Lemma 2.1, p.652-653 in Corradi (1999).  $II = o_{a.s}(1)$  because of the law of the iterated logarithm for strong mixing processes, that follows as a corollary of Theorem 1 in Eberlain.

Thus it suffices to show that

$$\limsup_{T \rightarrow \infty} \sup_{r \in [\gamma, 1-\gamma]} \frac{1}{d_T} |B(Tr) - rB(T)| = \sigma_0 1/2 \quad (14)$$

as

$$\limsup_{T \rightarrow \infty} \sup_{r \in [\gamma, 1-\gamma]} d_T^{-1} \left| \sum_{t=1}^{[Tr]} \epsilon_t - r \sum_{t=1}^T \epsilon_t \right| = \limsup_{T \rightarrow \infty} \sup_{r \in [\gamma, 1-\gamma]} d_T^{-1} |B(Tr) - rB(T)| + o_{a.s}(1)$$

In order to show (14), we shall proceed in three steps.

STEP 1: Show that pointwise in  $r$ ,  $\limsup_{T \rightarrow \infty} d_T^{-1} |B(Tr) - rB(T)| = \sigma_0 \sqrt{r(1-r)}$ .

STEP 2: Show that  $d_T^{-1} |B(Tr) - rB(T)|$  is almost sure asymptotically uniform equicontinuous (a.s. AUEC) in  $r$  over  $[\gamma, 1-\gamma]$ .

STEP 3. Given Step 1 and Step 2, show that  $\limsup_{T \rightarrow \infty} \sup_{r \in [\gamma, 1-\gamma]} d_T^{-1} |B(Tr) - rB(T)| = \sigma_0 1/2$ .

PROOF of STEP 1: For any given  $r$ ,

$$\begin{aligned} & \limsup_{T \rightarrow \infty} \sup_{r \in [\gamma, 1-\gamma]} \frac{1}{d_T} |B(Tr) - rB(T)| \\ &= \limsup_{T \rightarrow \infty} \sup_{r \in [\gamma, 1-\gamma]} d_T^{-1} (1-r) \left| Z\left(\frac{r}{1-r}T\right) \right| = \sigma_0 \sqrt{r(1-r)} \end{aligned} \quad (15)$$

with  $Z$  be a Brownian motion with variance equal to  $\sigma_0^2$ . The first equality in (15) comes from the fact that  $B(Tr) - rB(T)$  is a Brownian Bridge and has the same distribution of  $(1-r)Z(\frac{r}{1-r}T)$  (see e.g. Sen, 1986, p.113). The last equality in (15) follows from the LIL for the Brownian motion (e.g. Karatzas and Shreve, 1991, p.112), for any given  $r$ .

PROOF OF STEP 2:

$$d_T^{-1} \left| (B(Tr) - rB(T)) - (B(Tr') - r'B(T)) \right|$$

$$\leq d_T^{-1} |B(Tr) - B(Tr')| + d_T^{-1} |(r - r')B(T)|$$

Let  $S(r, \delta) = (r' : |r - r'| \leq \delta)$ . Now, because of the LIL for the Brownian motion, almost surely

$$\limsup_{T \rightarrow \infty} \sup_{r \in [\gamma, 1-\gamma]} \sup_{r' \in S(r, \delta)} d_T^{-1} |(r - r')B(T)| \leq 2\delta\sigma_0$$

and it approaches 0 as  $\delta \rightarrow 0$ , ensuring the a.s.AUEC of the term.

For given  $T$ , because of the Holder continuity of order  $\eta \in (0, 1/2)$  of the Brownian motion (see eg. Karatsas and Shreve, p.53-54), we have that almost surely,

$$\sup_{r \in [\gamma, 1-\gamma]} \sup_{r' \in S(r, \delta)} \frac{|B(r) - B(r')|}{|r - r'|^\eta} < C$$

where  $C$  is an appropriate constant. By noting that  $d_T^{-1} \rightarrow 0$  as  $T \rightarrow \infty$ , it follows that almost surely, as  $\delta \rightarrow 0$ ,

$$\limsup_{T \rightarrow \infty} \sup_{r \in [\gamma, 1-\gamma]} \sup_{r' \in S(r, \delta)} d_T^{-1} |B(r) - B(r')| = 0$$

PROOF OF STEP 3: Let  $S(\delta, r_i) = (r : |r - r_i| \leq \delta)$ , so that we cover the space  $[\gamma, 1 - \gamma]$  with  $N$  balls of radius  $\delta$ . Now for  $r \in S(\delta, r_i)$ ,

$$\begin{aligned} d_T^{-1} |B(Tr) - rB(T)| &\leq \sup_{r' \in S(\delta, r_i)} d_T^{-1} |B(Tr') - r'B(T)| \\ &\leq \sup_{r' \in S(\delta, r_i)} d_T^{-1} |(B(Tr') - r'B(T)) - (B(Tr_i) - r_iB(T))| \\ &\quad + d_T^{-1} |B(Tr_i) - r_iB(T)| \end{aligned}$$

and so

$$\begin{aligned} &\sup_{r' \in [\gamma, 1-\gamma]} d_T^{-1} |B(Tr) - rB(T)| \\ &\leq \sup_{r' \in [\gamma, 1-\gamma]} \sup_{r' \in S(\delta, r_i)} d_T^{-1} |(B(Tr') - r'B(T)) - (B(Tr_i) - r_iB(T))| \quad (16) \\ &\quad + \max_{i=1, \dots, N} d_T^{-1} |B(Tr_i) - r_iB(T)| \end{aligned}$$

From Step 2, we know that as  $\delta \rightarrow 0$ , the lim sup as  $T \rightarrow \infty$  of the first term in the RHS of (16) is almost surely zero. Now, from the proof of Step 1,

$$\begin{aligned} &\limsup_{T \rightarrow \infty} \max_{i=1, \dots, N} d_T^{-1} |B(Tr_i) - r_iB(T)| \\ &= \limsup_{T \rightarrow \infty} \max_{i=1, \dots, N} d_T^{-1} |(1 - r_i)Z\left(\frac{r_i}{1 - r_i}T\right)| = \sigma_0 1/2 \end{aligned}$$

given (15). This concludes the proof of (a).

(b)  $\hat{\sigma}_T^2 \xrightarrow{a.s.} \sigma_0^2$  from Theorem 1 in de Jong (1999), once we have shown that his Assumptions (1)-(4) hold. (1) and (4) hold straightforwardly, given the Bartlett kernel we are using. Given A1 and A2, de Jong (2) holds (using his notation) for  $r = 4$  and  $p = 2(2 + \delta)$  and (3) holds given the rate of growth for the lag truncation parameter in the statement of the proposition. This concludes the proof of (i).

(ii) First note that under A1 – A2, by the strong law of large numbers (e.g. McLeisch 1975, Theorem 2.10),

$$\begin{aligned} \bar{X} &= \frac{1}{T} \sum_{t=1}^T X_t = \frac{k_1}{T} \frac{1}{k_1} \sum_{t=1}^{k_1} X_t + \frac{k_2 - k_1}{T} \frac{1}{k_2 - k_1} \sum_{t=k_1+1}^{k_2} X_t + \dots \\ &\quad + \frac{T - k_m}{T} \frac{1}{T - k_m} \sum_{t=k_{m+1}+1}^T X_t \xrightarrow{a.s.} \sum_{i=1}^{m+1} (\tau_i - \tau_{i-1}) \mu_i = \mu \end{aligned}$$

We shall now show that  $\frac{\hat{\sigma}_{1T}^2}{l_T} = O_{as}(1)$ . First

$$\frac{1}{Tl_T} \sum_{t=1}^T (X_t - \bar{X})^2 = \frac{1}{Tl_T} \sum_{t=1}^T X_t^2 + \frac{1}{l_T} \bar{X}^2 - 2\bar{X} \frac{1}{Tl_T} \sum_{t=1}^T X_t = o_{as}(1)$$

by the strong law of large numbers, and by recalling that  $\frac{\sqrt{T}}{\sqrt{2 \log(\log(T))}} \bar{X} = O_{a.s.}(1)$ . Second,

$$\begin{aligned} &\frac{1}{Tl_T} \sum_{s=1}^{l_T} w_s \sum_{t=s+1}^T (X_t - \bar{X})(X_{t-s} - \bar{X}) \\ &= \frac{1}{Tl_T} \sum_{s=1}^{l_T} w_s \sum_{t=s+1}^T (X_t - E(X_t))(X_{t-s} - E(X_{t-s})) + \frac{1}{Tl_T} \sum_{s=1}^{l_T} w_s \sum_{t=s+1}^T (X_t - E(X_t))(E(X_{t-s}) - \bar{X}) \\ &\quad + \frac{1}{Tl_T} \sum_{s=1}^{l_T} w_s \sum_{t=s+1}^T (X_{t-s} - E(X_{t-s}))(E(X_t) - \bar{X}) - \frac{1}{Tl_T} \sum_{s=1}^{l_T} w_s \sum_{t=s+1}^T (\bar{X} - E(X_t))(\bar{X} - E(X_{t-s})) \\ &= o_{as}(1) + O_{as}(1) \end{aligned}$$

by noting that the first three terms on the RHS are  $o_{as}(1)$  by the strong law of large number and as  $\sup_t E(X_t) < \infty$ . The last term is  $O_{a.s.}(1)$ , as  $\bar{X} \xrightarrow{a.s.} \sum_{i=1}^{m+1} (\tau_i - \tau_{i-1}) \mu_i = \mu$

As for the numerator, for  $r$ , such that  $[Tr] = k_1$ , it can be rewritten as

$$\begin{aligned} & d_T^{-1} \left| \sum_{j=1}^{k_1} (X_j - E(X_j)) - (\bar{X} - \mu) + (E(X_j) - \mu) \right| \\ & \geq d_T^{-1} \left| \sum_{j=1}^{k_1} (E(X_j) - \mu) \right| - d_T^{-1} \left| \sum_{j=1}^{k_1} (X_j - E(X_j)) - k_1(\bar{X} - \mu) \right| = A_T + C_T \\ & C_T \leq d_T^{-1} \left| \sum_{j=1}^{k_1} (X_j - E(X_j)) \right| + \frac{k_1}{\sqrt{2T \log \log T}} (\bar{X} - \mu) = O_{as}(1) \end{aligned}$$

because of the LIL for strong mixing processes, given that  $\gamma T \leq k_1 \leq (1 - \gamma)T$ .

$$A_T = |\mu_0 - \mu| \frac{k_1}{d_T} \geq \Delta \gamma \frac{T}{\sqrt{2T \log \log T}}$$

because of A4 and A5. As  $d_T^{-1} \sup_{r \in [\gamma, 1-\gamma]} \left| \sum_{j=1}^{[Tr]} (X_j - \bar{X}) \right| \geq d_T^{-1} \left| \sum_{j=1}^{k_1} (X_j - \bar{X}) \right|$ , the desired result follows. ■

PROOF OF PROPOSITION 3.2.

(i) Given that  $T^\xi \rightarrow \infty$ , as  $T \rightarrow \infty$ ,  $\forall t$ ,  $\bar{X}^t - \bar{X} \xrightarrow{a.s.} 0$ , the result follows by the same argument used in the proof of (i) in Proposition 3.1.

(ii) It suffices to show that, under  $H_A$ ,  $\sigma_{L,T}^2 \xrightarrow{a.s.} \sigma_0^2$

$$\begin{aligned} \sigma_{L,T}^2 &= \frac{1}{T} \sum_{t=1}^T ((X_t - \mu_0 \mathbf{1}\{1, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\}) - (\bar{X}^t - \mu_0 \mathbf{1}\{1, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\}))^2 + \\ & \quad \frac{2}{T} \sum_{s=1}^{l_T} w_s \sum_{t=s+1}^T ((X_t - \mu_0 \mathbf{1}\{1, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\}) - (\bar{X}^t - \mu_0 \mathbf{1}\{1, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\})) \times \\ & \quad ((X_{t-s} - \mu_0 \mathbf{1}\{1, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\}) - (\bar{X}^{t-s} - \mu_0 \mathbf{1}\{1, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\})) \\ &= I + II + III + IV + V + VI \end{aligned}$$

where

$$\begin{aligned} I &= \frac{1}{T} \sum_{t=1}^T (X_t - \mu_0 \mathbf{1}\{1, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\})^2 + \\ & \frac{2}{T} \sum_{s=1}^{l_T} w_s \sum_{t=s+1}^T (X_t - \mu_0 \mathbf{1}\{1, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\}) \times (X_{t-s} - \mu_0 \mathbf{1}\{1, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\}) \end{aligned}$$



$$\begin{aligned}
II &= -\frac{1}{T} \sum_{t=1}^T (\bar{X}^t - \mu_0 \mathbf{1}\{1, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\})^2 \\
III &= -\frac{2}{T} \sum_{t=1}^T (\bar{X}^t - \mu_0 \mathbf{1}\{1, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\}) \times (X_t - \mu_0 \mathbf{1}\{1, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\}) \\
IV &= \frac{2}{T} \sum_{s=1}^{l_T} w_\tau \sum_{t=s+1}^T (\bar{X}^t - \mu_0 \mathbf{1}\{1, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\}) \times (\bar{X}^{t-s} - \mu_0 \mathbf{1}\{s, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\}) \\
V &= \frac{2}{T} \sum_{s=1}^{l_T} w_s \sum_{t=s+1}^T (X_t - \mu_0 \mathbf{1}\{1, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\}) \times (\bar{X}^{t-s} - \mu_0 \mathbf{1}\{s, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\}) \\
VI &= \frac{2}{T} \sum_{s=1}^{l_T} w_s \sum_{t=s+1}^T (X_{t-s} - \mu_0 \mathbf{1}\{s, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\}) \times (\bar{X}^t - \mu_0 \mathbf{1}\{s, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\})
\end{aligned}$$

Now  $I \xrightarrow{a.s.} \sigma_0^2$  by Theorem 1 in de Jong (1999). Thus we need to show that  $II, III, IV, V, VI$  are  $o_{a.s.}(1)$ . We shall begin by showing that  $VI$  is  $o_{a.s.}(1)$ . We can rewrite  $VI$  as

$$\begin{aligned}
VI &= \frac{2}{T} \sum_{s=1}^{l_T} w_\tau \sum_{t=s+1}^{k_1 - T^\xi} (X_{t-s} - \mu_0) (\bar{X}^t - \mu_0) + \frac{2}{T} \sum_{s=1}^{l_T} w_s \sum_{t=k_1 + T^\xi + s}^T (X_{t-s} - \mu_1) (\bar{X}^t - \mu_1) \\
&+ \frac{2}{T} \sum_{s=1}^{l_T} w_\tau \sum_{t=k_1 - T^\xi + 1}^{k_1 - 1 + T^\xi + s} (X_{t-s} - \mu_0 \mathbf{1}\{1, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\}) \times (\bar{X}^t - \mu_0 \mathbf{1}\{1, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\}) \\
&= VIa + VIb + VIc
\end{aligned}$$

Now  $\forall t \leq k_1$ ,  $\bar{X}^t - \mu_0 = O_{a.s.}(T^{-\xi/2} \sqrt{\log \log T^\xi})$  uniformly in  $t$ , and  $\forall t > k_1 + 1$ ,  $\bar{X}^t - \mu_1 = O_{a.s.}(T^{-\xi/2} \sqrt{\log \log T^\xi})$ .  $\frac{1}{T} \sum_{s=1}^{l_T} w_s \sum_{t=s+1}^{k_1 - 2T^\xi} (X_{t-s} - \mu_0) = o_{a.s.}(l_T)$ , thus as  $\frac{T^\xi}{l_T^2 \log \log T^\xi} \rightarrow \infty$  as  $T \rightarrow \infty$ ,  $VIa$  is  $o_{as}(1)$ , and for the same argument  $VIb$  is  $o_{a.s.}(1)$  too. Finally, as for  $VIc$ ,  $(\bar{X}^t - \mu_0 \mathbf{1}\{1, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\}) = O_{a.s.}(1)$ , as  $\bar{X}^t$  satisfy a strong law of large numbers, also

$$\frac{1}{T} \sum_{s=1}^{l_T} w_s \sum_{t=k_1 - T^\xi + 1}^{k_1 + 1 + T^\xi - 1 + s} (X_{t-s} - \mu_0 \mathbf{1}\{1, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\}) = O_{a.s.}(l_T T^{\xi-1} \log \log T^\xi)$$

Thus for  $\frac{l_T T^\xi \log \log T^\xi}{T} \rightarrow 0$ ,  $VIc$  is  $o_{a.s.}(1)$  too. Now,  $V$  is  $o_{a.s.}(1)$  by the same argument. As for  $IV$  it can be written as

$$IV = \frac{2}{T} \sum_{s=1}^{l_T} w_s \sum_{t=k_1 + T^\xi + s}^T (\bar{X}^t - \mu_1) \times (\bar{X}^{t-s} - \mu_1) + \frac{2}{T} \sum_{s=1}^{l_T} w_s \sum_{t=s+1}^{k_1 - T^\xi} (\bar{X}^t - \mu_0) \times (\bar{X}^{t-s} - \mu_0)$$

$$\begin{aligned}
& + \frac{2}{T} \sum_{s=1}^{l_T} w_s \sum_{t=k_1+1-T^\xi}^{k_1+T^\xi-1+s} (\bar{X}^t - \mu_0 \mathbf{1}\{1, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\}) \\
& \times (\bar{X}^{t-s} - \mu_0 \mathbf{1}\{s, k_1\} - \mu_1 \mathbf{1}\{k_1 + 1, T\}) \\
& = IVa + IVb + IVc
\end{aligned}$$

$IVa$  and  $IVb$  are  $O_{a.s.}(l_T T^{-\xi} \log \log T)$  and  $IVc$  is  $O_{a.s.}(l_T T^{\xi-1} \log \log T)$ , thus  $IV$  is  $o_{a.s.}(1)$ . Finally  $II$  and  $III$  are  $o_{a.s.}(1)$  by the same argument used above. ■

### PROOF OF LEMMA 3.3

Recall that that we can write  $S_T(k)$ , as

$$\frac{S_T(k)}{T} = \frac{\sum_{t=1}^T x_t^2}{T} - \frac{k}{T-k} \mu \left(\frac{k}{T}\right)^2 - \frac{T}{T-k} \bar{\mu}^2 + 2 \frac{k}{T-k} \bar{\mu} \mu \left(\frac{k}{T}\right).$$

where  $\mu$  and  $\hat{\mu}$  are defined as below equation (5). The first term of  $\frac{S_T(k)}{T}$  is

$$\frac{\sum_{t=1}^T x_t^2}{T} = \frac{\sum_{t=1}^T \varepsilon_t^2}{T} + \sum_{i=1}^{m_0+1} \mu_i^2 \left( \frac{[\tau_i T]}{T} - \frac{[\tau_{i-1} T]}{T} \right) + \frac{2}{T} \sum_{i=1}^{m_0+1} \mu_i \sum_{t=[\tau_{i-1} T]+1}^{[\tau_i T]} \varepsilon_t$$

the first and the last term on the RHS converge almost surely to  $\sigma^2$  and to zero by a strong law of large number (SLLN) (e.g. McLeish 1975), while the remaining term is deterministic and goes to  $\sum_{i=1}^m \mu_i^2 (\tau_i - \tau_{i-1})$ . The sample mean  $\bar{\mu}$  converges to  $\mu$  by SLLN. The mean in the first subsample  $\mu(\frac{[\tau T]}{T})$  converges a.s. uniformly in  $\tau$  to  $\mu(\tau)$  given that for every  $\tau$  SLLN ensures the convergence, also a strongly stochastically equicontinuity conditions (Andrews, 1992) is satisfied; the same holds for  $\mu(\frac{[\tau T]}{T})^2$ . Because all the stochastic terms have uniform limit on  $\tau \in [\gamma, 1 - \gamma]$ , the result then follows. ■

The proof of Theorem 3.4 requires three Lemmas.

#### LEMMA A1

There exists a constant  $C$  depending only on the break points such that

$$U(\tau) - U(\tau_0^1) \geq C |\tau - \tau_0^1|$$

where  $\tau_0^1 = \arg \min(U(\tau_1), \dots, U(\tau_{m_0}))$ , so  $\tau_0^1$  is the break among the  $m_0$  that most contributes to the reduction of the sum of squared residuals.

PROOF: Consider a  $\tau \in [\tau_i, \tau_{i+1}]$  where  $\tau_i$  is a generic break point among the possible  $m_0$  and assume without loss of generality that  $U(\tau_{i+1}) > U(\tau_i)$ . The expression for  $U(\tau)$  gives that

$$\begin{aligned} U(\tau) - U(\tau_i) &= -\frac{\tau}{1-\tau}\mu(\tau)^2 - \frac{1}{1-\tau}\mu^2 + 2\frac{\tau}{1-\tau}\mu\mu(\tau) \\ &\quad + \frac{\tau_i}{1-\tau_i}\mu(\tau_i)^2 + \frac{1}{1-\tau_i}\mu^2 - 2\frac{\tau_i}{1-\tau_i}\mu\mu(\tau_i), \end{aligned}$$

and given that  $\mu(\tau) = \frac{(\tau-\tau_i)}{\tau}\mu_{i+1} + \frac{\tau_i}{\tau}\mu(\tau_i)$  and after some algebra it follows that

$$\begin{aligned} &U(\tau) - U(\tau_i) \\ &= \frac{(\tau - \tau_i)}{(1 - \tau)} \left\{ \frac{\tau_i}{\tau} (\mu(\tau_i) - \mu_{i+1})^2 - (\mu_{i+1} - \mu)^2 - \frac{\tau_i}{1 - \tau_i} (\mu(\tau_i) - \mu)^2 \right\} \\ &= (\tau - \tau_i) \left[ \frac{(1 - \tau_{i+1}) \tau_{i+1}}{\tau (1 - \tau) (\tau_{i+1} - \tau_i)} \right] \\ &\quad \frac{(\tau_{i+1} - \tau_i)}{(1 - \tau_{i+1})} \left\{ \frac{\tau_i}{\tau_{i+1}} (\mu(\tau_i) - \mu_{i+1})^2 - \frac{\tau}{\tau_{i+1}} (\mu_{i+1} - \mu)^2 - \frac{\tau}{\tau_{i+1}} \frac{\tau_i}{1 - \tau_i} (\mu(\tau_i) - \mu)^2 \right\} \\ &\geq (\tau - \tau_i) \left[ \frac{(1 - \tau_{i+1}) \tau_{i+1}}{\tau (1 - \tau) (\tau_{i+1} - \tau_i)} \right] \\ &\quad \frac{(\tau_{i+1} - \tau_i)}{(1 - \tau_{i+1})} \left\{ \frac{\tau_i}{\tau_{i+1}} (\mu(\tau_i) - \mu_{i+1})^2 - (\mu_{i+1} - \mu)^2 - \frac{\tau_i}{1 - \tau_i} (\mu(\tau_i) - \mu)^2 \right\} \\ &= (\tau - \tau_i) \left[ \frac{(1 - \tau_{i+1}) \tau_{i+1}}{\tau (1 - \tau) (\tau_{i+1} - \tau_i)} \right] \{U(\tau_{i+1}) - U(\tau_i)\} \\ &\geq C_i (\tau - \tau_i) \end{aligned}$$

the second equality is obtained just arranging terms, the first inequality is due to  $\tau \leq \tau_{i+1}$ , the third equality follows by the definition of  $U(\tau_{i+1}) - U(\tau_i)$  and finally the last inequality is obtained by substituting to the expression into brackets the positive constant  $C_i$  which is only function of breaks fraction  $\tau_{i+1}$  and  $\tau_i$  and the difference of the sum of squared residuals at the break dates. Given that the number of break points is finite and given A4 then there is a constant  $c$  such that

$$U(\tau_i) - U(\tau_0^1) \geq c |\tau_i - \tau_0^1|$$

and combining this with the previous finding

$$\begin{aligned} U(\tau) - U(\tau_0^1) &\geq C_i |\tau - \tau_i| + c |\tau_i - \tau_0^1| \\ &\geq C |\tau - \tau_0^1| \end{aligned}$$

where  $C$  is the min of  $c$  and the finitely many  $C_i$ .

LEMMA A2

Under assumption A1–A2 and A4,  $|\hat{\tau}^1 - \tau_0^1| = O_{as}(T^{-1/2}\sqrt{\log \log T})$ .

PROOF: Let

$$\begin{aligned} \frac{S_T(\hat{k})}{T} - \frac{S_T([\tau^0 T])}{T} &= \frac{S_T(\hat{k})}{T} - \frac{S_T([\tau^0 T])}{T} \pm U(\hat{\tau}^1) \pm U(\tau_0^1) \\ &\geq -2 \sup_k \left| \frac{S_T(k)}{T} - U(k/T) \right| + U(\hat{\tau}^1) - U(\tau_0^1) \end{aligned}$$

By the definition of the break estimator the left hand side of the above inequality is less than zero and so we obtain that

$$2 \sup_k \left| \frac{S_T(k)}{T} - U(k/T) \right| \geq U(\hat{\tau}^1) - U(\tau_0^1) \geq C |\hat{\tau}^1 - \tau_0^1|$$

where the last inequality follows by Lemma A1. Given that the left most term converges to zero almost surely at rate  $T^{-1/2}\sqrt{\log \log T}$  by the same argument used in the proof of Proposition 3.1 (i), the desired result follows.

LEMMA A3

Under assumption A1–A2 and A4, there is an  $M_T = O(T^{-\eta})$ ,  $\eta > 0$  such that

$$\min_{k \in D_T} (S_T(k) - S_T([\tau_0^1 T])) \geq 0 \quad a.s.$$

where  $D_T = \{k : T\tau_0^i(1 + \rho) \leq k \leq T\tau_0^{i+1}(1 - \rho), |k - [\tau_0^1 T]| > M_T\}$  with  $\tau_0^i$  and  $\tau_0^{i+1}$  being the two break points just before and after  $\tau_0^1$ , and  $\rho$  is a small positive constant.

PROOF: Rewrite for  $\tau = \frac{k}{T}$ ,  $\tau_0^1 = \frac{k_0^1}{T}$ , and assume, without loss of generality,  $k \leq k_0^1$ . Note that the same result will hold for the case of  $\tau = \lim_{T \rightarrow \infty} \frac{k}{T}$ , in fact the steps below would hold up to a term vanishing at rate  $T^{-1}$ . Consider a  $k$  for which it is true that  $S_T(k) \leq S_T([\tau_0^1 T])$ , then it follows that

$$\frac{S_T(k) - S_T([\tau_0^1 T]) - T(U(\tau) - U(\tau_0^1))}{|k - [\tau_0^1 T]|} \leq -C, \quad (17)$$

where  $C$  is the constant of Lemma A1. Note that the numerator on the left hand side is negative. The set  $D_T$  represent the data points which are bounded away from the previous and

the following break dates and which are distant from the date of the break  $\tau_0^1$  of at least  $M_T$  observations.

The proof is based on arguing that with probability one for all  $k \in D_T$  the absolute value of (17) is zero and so by contradiction we have that  $S_T(k) \geq S_T([\tau_0^1 T])$  for  $k \in D_T$ . Consider the numerator of the previous expression and indicate with  $k_0^1 = [\tau_0^1 T]$

$$\begin{aligned} & |S_T(k) - S_T(k_0^1) - T(U(\tau) - U(\tau_0^1))| = \\ & T \left| -\frac{k}{T-k} \mu\left(\frac{k}{T}\right)^2 - \frac{T}{T-k} \bar{\mu}^2 + 2\frac{k}{T-k} \bar{\mu} \mu\left(\frac{k}{T}\right) \right. \\ & \left. + \frac{k_0^1}{T-k_0^1} \mu\left(\frac{k_0^1}{T}\right)^2 + \frac{T}{T-k_0^1} \bar{\mu}^2 - 2\frac{k_0^1}{T-k_0^1} \bar{\mu} \mu\left(\frac{k_0^1}{T}\right) \right. \\ & \left. + \frac{\tau}{1-\tau} \mu(\tau)^2 + \frac{1}{1-\tau} \mu^2 - 2\frac{\tau}{1-\tau} \mu \mu(\tau) \right. \\ & \left. - \frac{\tau_0^1}{1-\tau_0^1} \mu(\tau_0^1)^2 - \frac{1}{1-\tau_0^1} \mu^2 + 2\frac{\tau_0^1}{1-\tau_0^1} \mu \mu(\tau_0^1) \right| \end{aligned}$$

The proof is based on showing that the left hand side of the above expression divided by  $|k - k_0^1|$  goes to zeros almost surely. By triangular inequality, it follows that

$$\begin{aligned} & |S_T(k) - S_T(k_0^1) - T(U(\tau) + U(\tau_0^1))| \leq \\ & T \left| -\frac{k}{T-k} \mu\left(\frac{k}{T}\right)^2 + \frac{k_0^1}{T-k_0^1} \mu\left(\frac{k_0^1}{T}\right)^2 + \frac{\tau}{1-\tau} \mu(\tau)^2 - \frac{\tau_0^1}{1-\tau_0^1} \mu(\tau_0^1)^2 \right| + \\ & T \left| -\frac{T}{T-k} \bar{\mu}^2 + \frac{T}{T-k_0^1} \bar{\mu}^2 + \frac{1}{1-\tau} \mu^2 - \frac{1}{1-\tau_0^1} \mu^2 \right| + \\ & 2T \left| \frac{k}{T-k} \bar{\mu} \mu\left(\frac{k}{T}\right) - \frac{k_0^1}{T-k_0^1} \bar{\mu} \mu\left(\frac{k_0^1}{T}\right) - \frac{\tau}{1-\tau} \mu \mu(\tau) + \frac{\tau_0^1}{1-\tau_0^1} \mu \mu(\tau_0^1) \right|. \end{aligned}$$

We call the three elements on the right hand side as  $I1_T$ ,  $I2_T$  and  $I3_T$  respectively. The first term on the right hand side,  $I1_T$ , is equal to,

$$\begin{aligned} & T \left| -\frac{k}{T-k} \left( \frac{\sum^k (\mu_t + \varepsilon_t)}{k} \right)^2 + \frac{k_0^1}{T-k_0^1} \left( \frac{\sum^{k_0^1} (\mu_t + \varepsilon_t)}{k_0^1} \right)^2 + \frac{\tau}{1-\tau} \mu(\tau)^2 - \frac{\tau_0^1}{1-\tau_0^1} \mu(\tau_0^1)^2 \right| \leq \\ & T \left| -\frac{1}{k(T-k)} \left( \sum^k \mu_t \right)^2 - \frac{1}{k(T-k)} \left( \sum^k \varepsilon_t \right)^2 - \frac{2}{k(T-k)} \left( \sum^k \mu_t \sum^k \varepsilon_t \right) \right| \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{k_0^1 (T - k_0^1)} \left( \sum^{\binom{k_0^1}{k_0^1}} \mu_t \right)^2 + \frac{1}{k_0^1 (T - k_0^1)} \left( \sum^{\binom{k_0^1}{k_0^1}} \varepsilon_t \right)^2 + \frac{2}{k_0^1 (T - k_0^1)} \left( \sum^{\binom{k_0^1}{k_0^1}} \mu_t \sum^{\binom{k_0^1}{k_0^1}} \varepsilon_t \right) \\
& + \frac{\tau}{1 - \tau} \mu(\tau)^2 - \frac{\tau_0^1}{1 - \tau_0^1} \mu(\tau_0)^2 |,
\end{aligned}$$

where  $\mu_t = \mu_i 1\{k_i + 1, k_{i+1}\}$ ,  $\sum^k$  is the summation from 1 to  $k$ , while  $\sum_{k_0}^{k_0}$  is the summation from  $k$  to  $k_0$ . We will show that the right hand side normalized by  $|k - k_0^1|$  goes to zero. First consider the first and the seventh elements on the right hand side

$$\begin{aligned}
& T \left| \frac{k}{(T - k)} \left( \frac{\sum^k \mu_t}{k} \right)^2 - \frac{\tau}{1 - \tau} \mu(\tau)^2 \right| = \\
& \left| \frac{T\tau}{(1 - \tau)} \left( \frac{\sum^k \mu_t}{k} - \mu(\tau) \right) \left( \frac{\sum^k \mu_t}{k} + \mu(\tau) \right) \right|
\end{aligned}$$

which divided by  $|k - k_0^1|$  it converges to zero, as  $|k - k_0^1| > M_T$ ,  $M_T \rightarrow \infty$ , by the fact that  $(\frac{\sum^k \mu_t}{k} - \mu(\tau)) = O(T^{-1})$  and  $(\frac{\sum^k \mu_t}{k} + \mu(\tau)) = O(1)$ . The same result holds for the fourth and the last term. The second and the fifth element are

$$\begin{aligned}
& T \left| \frac{1}{k(T - k)} \left( \sum^k \varepsilon_t \right)^2 - \frac{1}{k_0^1 (T - k_0^1)} \left( \sum^{\binom{k_0^1}{k_0^1}} \varepsilon_t \right)^2 \right| = \\
& T \left| \frac{1}{k_0^1 (T - k_0^1)} \left[ \left( \sum^k \varepsilon_t \right)^2 - \left( \sum^{\binom{k_0^1}{k_0^1}} \varepsilon_t \right)^2 \right] + \frac{(k - k_0^1) (T - k - k_0^1)}{k(T - k) k_0^1 (T - k_0^1)} \left( \sum^k \varepsilon_t \right)^2 \right| = \\
& \left| \frac{T}{k_0^1 (T - k_0^1)} \left[ \left( \sum^{\binom{k_0^1}{k_0^1}} \varepsilon_t \right)^2 - 2 \sum^k \varepsilon_t \sum^{\binom{k_0^1}{k_0^1}} \varepsilon_t \right] + \frac{T(k - k_0^1) (T - k - k_0^1)}{k(T - k) k_0^1 (T - k_0^1)} \left( \sum^k \varepsilon_t \right)^2 \right|
\end{aligned}$$

all the terms are  $O_{as}(\log \log T)$  by the LIL and, divided by  $|k - k_0^1|$  which is  $O(T^{-\eta})$ ,  $\eta > 0$ , approach zero almost surely. Finally the third and sixth terms

$$\begin{aligned}
& T \left| \frac{1}{k(T - k)} \left( \sum^k \mu_t \sum^k \varepsilon_t \right) - \frac{1}{k_0^1 (T - k_0^1)} \left( \sum^{\binom{k_0^1}{k_0^1}} \mu_t \sum^{\binom{k_0^1}{k_0^1}} \varepsilon_t \right) \right| = \\
& T \left| \frac{1}{k_0^1 (T - k_0^1)} \left( \sum^k \mu_t \sum^k \varepsilon_t - \sum^{\binom{k_0^1}{k_0^1}} \mu_t \sum^{\binom{k_0^1}{k_0^1}} \varepsilon_t \right) - \frac{(k - k_0^1) (T - k - k_0^1)}{k(T - k) k_0^1 (T - k_0^1)} \left( \sum^k \mu_t \sum^k \varepsilon_t \right) \right| =
\end{aligned}$$

$$\left| \frac{T}{k_0^1 (T - k_0^1)} \left( \sum^k \mu_t \sum_k^{k_0^1} \varepsilon_t - \sum_k \mu_t \sum_k^{k_0^1} \varepsilon_t \right) - \frac{T (k - k_0^1) (T - k - k_0^1)}{k (T - k) k_0^1 (T - k_0^1)} \left( \sum^k \mu_t \sum^k \varepsilon_t \right) \right|$$

again dividing both sides by  $|k - k_0^1|$  it follows that it converges almost surely to zero by the same argument. This implies that  $I1_T$  normalized by  $|k - k_0^1|$  converges almost surely to zero; the same holds for  $I2_T$  and  $I3_T$ .

Given that the LHS of (17) converges almost surely to zero, the proof of the lemma follows by contradiction.

PROOF OF THEOREM 3.4

$$\begin{aligned} & \Pr \left( \lim_{T \rightarrow \infty} \left| \widehat{k}^1 - k_0^1 \right| \frac{1}{M_T} > 0 \right) \\ & \leq \Pr \left( \lim_{T \rightarrow \infty} \widehat{k}^1 \notin D_T \cup \left( \lim_{T \rightarrow \infty} \widehat{k}^1 \in D_T, \lim_{T \rightarrow \infty} \left| \widehat{k}^1 - k_0^1 \right| > M_T \right) \right) \\ & \leq \Pr \left( \min_{k \in D_T} (S_T(k) - S_T([\tau_0^1 T])) \leq 0 \right) = 0 \end{aligned}$$

This concludes the proof. ■

PROOF OF PROPOSITION 3.5: (i) Under the null, there is only one break, so let  $k_0^1 = k_1$ .

$$\begin{aligned} & d_T^{-1} \sup_{r \in [\gamma, 1-\gamma]} \left| \sum_{t=1}^{[rT]} (X_t - 1\{1, \widehat{k}^1\} \widehat{X}_1 - 1\{\widehat{k}^1 + 1, T\} \widehat{X}_2) \right| = \\ & d_T^{-1} \sup_{r \in [\gamma, 1-\gamma]} \left| \sum_{t=1}^{[rT]} (X_t - 1\{1, k_1\} \bar{X}_1 - 1\{k_1 + 1, T\} \bar{X}_2) \right| + o_{as}(1) \end{aligned}$$

given that  $\widehat{k}^1 - k_1 = O_{as}(T^\eta)$  and  $\widehat{X}_i - \bar{X}_i = O_{as}(T^{-1/2(1-\eta)} \sqrt{\log \log T})$ ,  $i = 1, 2$  with  $\bar{X}_1 = \frac{1}{k_1} \sum_{t=1}^{k_1} X_t = \frac{1}{k_1} \sum_{t=1}^{k_1} (\mu_1 + \epsilon_t)$  and  $\bar{X}_2 = \frac{1}{T-k_1} \sum_{t=k_1+1}^T X_t = \frac{1}{T-k_1} \sum_{t=k_1+1}^T (\mu_2 + \epsilon_t)$ , so we have:

$$\begin{aligned} & d_T^{-1} \sup_{r \in [\gamma, 1-\gamma]} \left| \sum_{t=1}^{[rT]} (X_t - 1\{1, k_1\} \bar{X}_1 - 1\{k_1 + 1, T\} \bar{X}_2) \right| + o_{as}(1) \\ & = d_T^{-1} \sup_{r \in [\gamma, 1-\gamma]} \left| \sum_{t=1}^{[rT]} \left( \epsilon_t - 1\{1, k_1\} \frac{\sum_{t=1}^{k_1} \epsilon_t}{k_1} - 1\{k_1 + 1, T\} \frac{\sum_{t=k_1+1}^T \epsilon_t}{T - k_1} \right) \right| + o_{as}(1) \\ & = d_T^{-1} \sup_{r \in [\gamma, 1-\gamma]} \left| \sum_{t=1}^{[rT]} \epsilon_t - 1\{1, k_1\} [rT] \frac{\sum_{t=1}^{k_1} \epsilon_t}{k_1} - 1\{k_1 + 1, T\} ([rT] - k_1) \frac{\sum_{t=k_1+1}^T \epsilon_t}{T - k_1} \right| + o_{as}(1) \end{aligned}$$

$$\begin{aligned}
&= d_T^{-1} \sup_{r \in [\gamma, 1-\gamma]} |1\{1, k_1\} \left( \sum_{t=1}^{[rT]} \epsilon_t - [rT] \frac{\sum_{t=1}^{k_1} \epsilon_t}{k_1} \right) + \\
&\quad 1\{k_1 + 1, T\} \left( \sum_{t=k_1+1}^{[rT]} \epsilon_t - ([rT] - k_1) \frac{\sum_{t=k_1+1}^T \epsilon_t}{T - k_1} \right)| + o_{as}(1) \\
&\leq d_T^{-1} \max \left\{ \sup_{r \in [\gamma, 1-\gamma]} |1\{1, k_1\} \left| \sum_{t=1}^{[rT]} \epsilon_t - [rT] \frac{\sum_{t=1}^{k_1} \epsilon_t}{k_1} \right|, \right. \\
&\quad \left. \sup_{r \in [\gamma, 1-\gamma]} |1\{k_1 + 1, T\} \left| \sum_{t=k_1+1}^{[rT]} \epsilon_t - ([rT] - k_1) \frac{\sum_{t=k_1+1}^T \epsilon_t}{T - k_1} \right| \right\} + o_{a.s.}(1)
\end{aligned}$$

Finally, given that each terms inside the max operator converges to a brownian bridge, the lim sup as  $T \rightarrow \infty$  of the right hand side of the last inequality is majorized by 1/2 by the same argument used in the proof of proposition 3.1(i).

(ii) By the same argument used in the proof of proposition 3.1 (ii). ■

#### PROOF OF LEMMA 3.6

Given the  $T^{-(1-\eta)}$  consistency of the estimated break points, the difference between the expression in (9) and the one corresponding expression evaluated at the true break points is  $o_{as}(1)$ . Thus the result follows by the same argument as in Lemma 3.3.

#### PROOF OF COROLLARY 3.7

Given A4, and given that  $\hat{k}^1/T - \tau_0^1 = O_{as}(T^{-(1-\eta)})$ , it follows by the same argument used in the proof of Theorem 3.4. In fact at the second step we are dealing with a series of  $n - 1$  unknow break and we detect the break among the remaining  $n - 1$  that most contributed to the reduction of the sum of square errors.

#### PROOF OF THEOREM 3.8

When we fail to reject the null, we stop; if we fail to reject  $H_0^m : m \text{ breaks}$ , we mean that we have performed the sequential procedure  $m$  times.

$$P(\lim_{T \rightarrow \infty} m_T > m_0) \leq \sum_{i=1}^m P(\lim_{T \rightarrow \infty} V_{ST}^i > 1/2 | H_0^i \text{ true}) = 0$$

$$P(\lim_{T \rightarrow \infty} m_T < m_0) = P(\limsup_{T \rightarrow \infty} V_{ST}^m \leq 1/2 | H_0^m \text{ false}) = 0.$$



## References

- [1] Ait-Sahalia Y. (1996a), Nonparametric pricing of interest rate derivatives securities, *Econometrica*, 64, 527-560.
- [2] Ait-Sahalia Y. (1996b), Testing continuous time models of the spot interest rate, *Review of Financial Studies*, 9, 385-426.
- [3] Andrews D.W.K. (1991), Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, 59, 817-858.
- [4] Andrews D.W.K. (1992), Generic uniform convergence, *Econometric Theory*, 8, 241-257.
- [5] Andrews D.W.K. and E.Zivot (1992), Further evidence on the great crash, *Journal of Business and Economic Statistics*, 10, 251-270.
- [6] Bai J. (1997), Estimating multiple breaks one at time, *Econometric Theory*, 13 315-352.
- [7] Bai J. (1999), Likelihood ratio tests for multiple structural changes, *Journal of Econometrics*, 91, 299-323.
- [8] Bai J. and P. Perron (1998), Estimating and testing linear models with multiple structural changes, *Econometrica* , 66, 47-78.
- [9] Banerjee A., R.L. Lumsdaine and J.H. Stock (1992), Recursive and sequential tests of the unit root and trend break hypothesis: theory and international evidence, *Journal of Business and Economic Statistics*, 10, 271-288.
- [10] Chu C.S.J., M. Stinchcombe & H. White (1996), Monitoring Structural Changes, *Econometrica*, 64, 1045-1066
- [11] Chu C.S.J., K. Hornik and C.M. Kuan (1995), The moving estimate test for parameter stability, *Econometric Theory*, 11, 699-720.
- [12] Corradi V. (1999), Deciding between I(0) and I(1) via FLIL-based bounds, *Econometric Theory*, 15, 643-663.

- [13] De Jong R.M. (1999), A strong consistency proof for heteroskedasticity and autocorrelation consistent covariance estimator, *Econometric Theory*, forthcoming.
- [14] Diebold, F.X. and A.Inoue (1999), Long memory and structural change, manuscript, North Carolina State University and NYU.
- [15] Eberlain E. (1996), On strong invariance principles under dependence assumptions, *Annals of Probability* , 14, 260-270.
- [16] Hendry D.F. (1997), The econometrics of macroeconomic forecasting, *Economic Journal*, 107, 1330-1357.
- [17] Hidalgo J. and P.M. Robinson (1996), Testing for structural change in a long memory environment, *Journal of Econometrics*, 70, 159-174.
- [18] Karatzsas J. and S.E. Shreve (1991), *Brownian motion and stochastic calculus*, Springer and Verlag, New York.
- [19] Leyborne S., T.C. Mills and P. Newbold (1998), Spurious rejection by Dickey-Fuller tests in the presence of a break under the null, *Journal of Econometrics*, 87, 191-203.
- [20] Lumsdaine R.L. and D.H. Papell (1997), Multiple trend breaks and the unit root hypothesis, *Review of Economics and Statistics*, 212-217
- [21] Liu J., S. Wu and J.V. Zidek (1997), On segmented multivariate regressions, *Statistica Sinica* , 7, 497-525.
- [22] McLeish D.L. (1975), A maximal inequality and dependent strong laws, *Annals of Probability* , 5, 829-839.
- [23] Nunes L.C., C.M. Kuan and P. Newbold (1995), Spurious breaks, *Econometric Theory* , 11, 736-749.
- [24] Nunes L.C., C.M. Kuan and P. Newbold (1996), Spurious number of breaks, *Economic Letters*, 50, 175-178.

- [25] Perron P. (1989) The great crash, the oil price shock and the unit root hypothesis, *Econometrica*, 57, 1361-1401.
- [26] Pesaran H. and A. Timmermann (1999) Model instability and choice of observation window, mimeo, UCSD and University of Cambridge
- [27] Pritsker M. (1997), Nonparametric density estimation and test of continuous time interest rate models, *Review of Financial Studies*, 11, 449-488.
- [28] Romer C.D. and D.H. Romer (1989) Does monetary policy matter? A new test in the spirit of Friedman and Schwartz. *NBER Macroeconomic Annals*, 121-170.
- [29] Rudebusch G.D. (1995) Federal Reserve interest rate targeting, rational expectations and the term structure, *Journal of Monetary Economics*, 36, 679
- [30] Sen P.K. (1986), Two LIL-type results for the Brownian Bridge, *Journal of Multivariate Analysis*, 19, 113-118.
- [31] Strassen V. (1964), An invariance principle for the law of the iterated logarithm, *Probability Theory and Related Fields* , 13, 211-226.
- [32] Vogelsang T.J. (1998), Testing for a shift in mean without having to estimate serial-correlation parameters, *Journal of Business, Economic Statistics*,16, 73-80.
- [33] Yao Y.C. (1988), Estimating the number of change-points via Schwartz's criterion, *Statistics and Probability Letters*, 6, 181-189.
- [34] Zivot E. and Andrews D.W.K. (1992), Further evidence on the great crash, the oil price and the unit root hypothesis, *Journal of Business and Economic Statistics*, 10, 251-270.