Electronic Journal of Statistics
Vol. 8 (2014) 931–964
ISSN: 1935-7524
DOI: 10.1214/14-EJS916

On distinguishing multiple changes in mean and long-range dependence using local Whittle estimation*

Changryong Baek†

Department of of Statistics, Sungkyunkwan University, Korea
e-mail: crbaek@skku.edu

and

Vladas Pipiras‡

Dept. of Statistics and Operations Research, UNC at Chapel Hill, USA
e-mail: pipiras@email.unc.edu

Abstract: It is well known that changes in mean superimposed by a short-range dependent series can be confused easily with long-range dependence. A procedure to distinguish the two phenomena is introduced. The proposed procedure is based on the local Whittle estimation of the long-range dependence parameter applied to the series after removing changes in mean, and comparing the results to those obtained through the available CUSUM-like approaches. According to the proposed procedure, for example, volatility series in finance seem more consistent with the changes-in-mean models whereas hydrology and telecommunication series are more in line with long-range dependence. As part of this work, a new method based on the local Whittle estimation to find the number of breaks is also introduced and its consistency is proved for the changes-in-mean models.

MSC 2010 subject classifications: Primary 62M10, 62G10; secondary 60G18.
Keywords and phrases: Hypothesis tests, size and power, short- and long-range dependence, changes in mean, local Whittle estimator, bootstrap.

Received January 2013.

Contents

*The authors would like to thank Mark Ramirez at Arizona State University for providing the updated series on the congressional approval in the US.
†The work of the first author was supported in part by the Basic Science Research Program from the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT & Future Planning (NRF-2014R1A1A1006025).
‡The work of the second author is supported in part by the NSA grant H98230-13-1-0220.

1. Introduction

Long-range dependence (LRD, in short), also known as long memory, refers to a second-order stationary time series $X = \{X_n\}_{n\in\mathbb{Z}}$ with the autocovariance function decaying at large lags as

$$\gamma_X(h) = \mathrm{Cov}(X_0, X_h) \sim Ch^{2d-1}, \quad \text{as } h \to \infty, \tag{1.1}$$

or the spectral density diverging at zero as

$$f_X(\omega) \sim c|\omega|^{-2d}, \quad \text{as } \omega \to 0. \tag{1.2}$$

Here, $d \in (0, 1/2)$ is the LRD parameter, and $C, c > 0$ are two constants. See, for example, Beran (1994), Doukhan, Oppenheim and Taqqu (2003), Palma (2007). The conditions (1.1) and (1.2) are not equivalent in general, though they both hold for most LRD models of interest. LRD is used to model real time series in many areas as diverse as telecommunications (Park and Willinger (2000)), economics and finance (Robinson (2003)), hydrology (Dmowska and Saltzman (1999)).

Note that, under (1.1),

$$\sum_{h=-\infty}^{\infty} |\gamma_X(h)| = \infty. \tag{1.3}$$

Short-range dependence (SRD, in short), on the other hand, refers to the case where the sum of the autocovariances is finite (i.e. $\sum_{h=-\infty}^{\infty} |\gamma_X(h)| < \infty$), for example, when the autocovariance decays exponentially fast; or the spectral density satisfies (1.2) with $d = 0$. For the latter reason, the value $d = 0$ is often associated with SRD, whereas again $d \in (0, 1/2)$ corresponds to LRD.

A characteristic feature of LRD series is apparent changes in the local mean level across a range of larger scales.[1] Conversely, a series composed of several changes in mean superimposed by a SRD series will exhibit LRD when using common estimators of the LRD parameter. This confusion between LRD and nonstationary models, such as the model involving changes in mean, has been documented well (e.g. Klemeš (1974), Roughan and Veitch (1999), Diebold and Inoue (2001), Mikosch and Stărică (2004), Mills (2007)) and has attracted a lot of attention in the literature (e.g. Kuan and Hsu (1998), Berkes et al. (2006), Bisaglia and Gerolimetto (2009), Qu (2011), Baek and Pipiras (2012)).

A number of statistical procedures started to emerge recently aiming at distinguishing LRD and nonstationary models. Jach and Kokoszka (2008) showed that wavelet-based, maximum likelihood tests for SRD are robust to the presence of nonstationarities. Ohanissian, Russell and Tsay (2008) used temporal aggregation to test against nonstationary models. Iacone (2010), McCloskey and Perron (2013) employ common estimators of the LRD parameter in the Fourier domain by selecting carefully the range of frequencies to be used in estimation.

Our work contributes to these efforts by focusing on LRD and changes in mean. The *changes-in-mean model* (*CM model*, for short) is given by

$$X_j = \mu + \sum_{r=1}^{R} \Delta_r 1_{\{k_r < j \le n\}} + \epsilon_j, \quad j = 1, \dots, n, \tag{1.4}$$

where $\{k_r, r = 1, ..., R\}$ are the unknown R break points, $\mu + \Delta_1 + \ldots + \Delta_r$ represents the mean in the r-th regime $(k_r, k_{r+1}]$ and $\{\epsilon_j\}$ is a SRD series. More detailed assumptions on $\{\epsilon_j\}$ will be made below. When necessary to indicate the number of break points R, we will also denote (1.4) as the CM-R model. Testing and estimation for the CM models (1.4) are well developed, notably in Bai (1997), Bai and Perron (1998); see also a nice paper by Robbins et al. (2011). In this paper, we are interested in understanding here whether these, now standard tools and approaches could be useful for our purpose at distinguishing the CM models (1.4) and LRD.

Some progress in this direction was made in Berkes et al. (2006), Baek and Pipiras (2012) and Yau and Davis (2012). Berkes et al. (2006) use standard CUSUM-based methods to devise tests for CM models against LRD. As noted in Baek and Pipiras (2012), however, these tests have very little power against LRD. A test with a much larger power, based on the local Whittle estimation, was suggested in Baek and Pipiras (2012) when testing for the CM model with $R = 1$ break against LRD (with an obvious extension to multiple breaks). But no suggestion was made as to how the test can be used to distinguish between CM models and LRD. Yau and Davis (2012) also focus on the CM-1 and LRD models, but use instead a full Whittle likelihood and a likelihood ratio test.

[1] For this reason, LRD is also known as the Joseph effect, a biblical reference to seven years of great abundance, followed by seven years of famine.

The approach suggested here follows Baek and Pipiras (2012) in using the local Whittle estimation of the LRD parameter, in conjunction with the methods to find breaks in CM models. The basic idea is the following. Modeling changes in mean involves two key components: estimation of the break points k_r in (1.4), and estimation of the number of breaks R. We shall focus on sequential procedures to find break points as in Bai (1997). In this case, selecting R can be thought as a stopping rule. The break points k_r for the CM model are commonly found by using the CUSUM statistic (Berkes et al. (2006)) or the LSE statistic (Bai (1997)). The two statistics are recalled below in Section 2, and differ by a multiplicative factor. Common *available stopping rules* are based on the sup-F test of Bai (1997), and the CUSUM-based test of Berkes et al. (2006), referred to as the BHKS test below. These various statistics and rules are not very different, and the resulting performances are comparable.

For the purpose of using these tools at distinguishing CM models and LRD, we suggest another stopping rule to select R for CM models. The rule is based on the local Whittle (LW) estimation of the LRD parameter (Robinson (1995a)). More specifically, when the break points are found sequentially for CM models, we suggest to estimate the LRD parameter in the series with the changes in mean removed and stop when the hypothesis of SRD ($d = 0$) cannot be rejected. We will refer to this as the *LW stopping rule* or the *LW test*. As for the available stopping rules, we show (Theorem 3.2 below) that under suitable assumptions, the LW stopping rule estimates consistently the number of breaks R. Moreover, for CM models, the finite sample performance when using the LW test is comparable to those when using the available sup-F and BHKS tests.

How does this help at distinguishing CM models and LRD? Under LRD, the number of breaks estimated by the LW test is shown to converge to $+\infty$ (Theorem 3.3 below). The same is expected for the available stopping rules, for example, this is implied by Berkes et al. (2006) for the BHKS test. However, the finite sample simulations suggest that the LW stopping rule performs *very differently* from the available rules when applied to a LRD series. As in Baek and Pipiras (2012), the LW test has a much larger power when testing for a CM model against LRD. In terms of the estimated number of breaks, this translates into that, for a *LRD series*, the number of breaks found by the LW stopping rule is now much larger than that found by the available stopping rules.

This discussion suggests a procedure for distinguishing CM models and LRD. Specifically, estimate a CM model by using the available stopping rules and the proposed LW rule. If the results are comparable, this suggests a CM model. If a much larger number of breaks is found using the LW stopping rule, this suggests a LRD model.

We formalize this idea through a test based on the difference between the estimated numbers of breaks where under the null of a CM model, the p-value is obtained using the bootstrap. Justifying the bootstrap approach goes beyond the scope of this work and will be addressed in the future (some recent works on the bootstrap and changes-in-mean models are mentioned in Section 4.4).

The proposed procedure has a number of attractive features. Note that comparable results not only point to a CM model, but also estimate it. Under a

CM model, this is a result of the consistency of the employed procedures – the consistency of the LW test is proved in this work. The difference in the behavior for LRD series is a result of different powers of the tests against LRD (which is also a consequence of the fact that only the LW test is an MLE-based test under the assumption of LRD).

The proposed procedure performs very well in simulations. Moreover, the results are also very encouraging when the procedure is applied to real data exhibiting LRD. Several series of stock price volatility are found to be more consistent with CM models. Several series from telecommunications and hydrology are found to be more in line with LRD. The same conclusions were reached by McCloskey and Perron (2013) using a different approach, and likely conform to the viewpoint held by many researchers working in the area. In a perhaps less known application of LRD series, the procedure is also applied to a series representing the congressional approval in the US.

The structure of the paper is as follows. We start by reviewing popular methods to find breaks, namely the least squares method (LSE) and the so-called CUSUM method in Section 2. The available stopping rules are discussed in Section 3. Also, our proposed method based on the local Whittle estimation is introduced and its consistency result for CM models and the divergence result for LRD models are stated. The method is examined through simulations, and the procedure to compare the number of estimated breaks is introduced in Section 4. Applications to several real data sets can be found in Section 5. Concluding remarks are in Section 6. Proofs of the consistency and divergence results are provided in Appendix.

2. Finding break points

Finding break points and change-point analysis in general have been an active research area over the past few decades. See, for example, Csörgő and Horváth (1997), Perron (2006) for comprehensive reviews. Here, we briefly review two popular methods to find break points and discuss their asymptotic results. We focus on linear sequences $\{\epsilon_j\}$ in the CM-R model (1.4) given by

$$\epsilon_j = \sum_{i=0}^{\infty} \psi_i Z_{j-i}, \quad \sum_{i=0}^{\infty} i|\psi_i| < \infty, \quad \sum_{i=0}^{\infty} \psi_i \neq 0, \tag{2.1}$$

where $\{Z_j\}$ are martingale differences satisfying $E(Z_j|\mathcal{F}_{j-1}) = 0$ with $\mathcal{F}_j = \sigma\{Z_i, i \leq j\}$, $EZ_j^2 = \sigma_Z^2$ and $\sup_j E|Z_j|^{2+\delta} < \infty$ for some $\delta > 0$.

2.1. LSE method

Assume first that there is only one break $R = 1$ at time k. Then, the CM-1 model becomes

$$X_j = \begin{cases} \mu + \epsilon_j, & j = 1, \dots, k, \\ \mu + \Delta_1 + \epsilon_j, & j = k+1, \dots, n. \end{cases} \tag{2.2}$$

The break time k can be estimated by minimizing the sum of squared residuals

$$\widehat{k}^L = \underset{1\le k\le n-1}{\operatorname{argmin}} \left(\sum_{j=1}^{k} (X_j - \overline{X}_k)^2 + \sum_{j=k+1}^{n} (X_j - \overline{X}_k^*)^2 \right), \tag{2.3}$$

where

$$\overline{X}_k = \frac{1}{k} \sum_{j=1}^{k} X_j, \quad \overline{X}_k^* = \frac{1}{n-k} \sum_{j=k+1}^{n} X_j \tag{2.4}$$

estimate μ and $\mu + \Delta_1$ respectively. We refer to (2.3) as the *least squares estimation* or the *LSE method*.

For later reference, observe also that

$$\widehat{k}^L = \underset{1\le k\le n-1}{\operatorname{argmax}} \left\{ \sum_{j=1}^{n} (X_j - \overline{X}_n)^2 - \left(\sum_{j=1}^{k} (X_j - \overline{X}_k)^2 + \sum_{j=k+1}^{n} (X_j - \overline{X}_k^*)^2 \right) \right\} \tag{2.5}$$

since $\sum_{j=1}^{n} (X_j - \overline{X}_n)^2$ does not depends on k. It can be seen further that (2.5) is equivalent to

$$\widehat{k}^L = \underset{1\le k\le n-1}{\operatorname{argmax}} \left| \left(\frac{k}{n} \left(1 - \frac{k}{n} \right) \right)^{1/2} \sqrt{n} \left(\overline{X}_k - \overline{X}_k^* \right) \right|. \tag{2.6}$$

When the errors $\{\epsilon_j\}$ are i.i.d. Gaussian, $\widehat{k}^L$ is also the maximum likelihood estimator (Csörgő and Horváth (1997), Section 1.6).

Multiple breaks for the CM-R model can be found by applying the LSE method in a recursive way. For example, after finding $\widehat{k}^L$ in (2.3), the LSE method is applied to the subsamples $\{X_1, \ldots, X_{\widehat{k}^L}\}$ and $\{X_{\widehat{k}^L+1}, \ldots, X_n\}$, and so on. Stopping rules for the method are discussed in Section 3.

Asymptotic properties of the LSE break estimators were studied by Bai (1997). Here, we briefly recall Bai's results in two directions. In one direction, the shifts Δ_i are *fixed* as the sample size n goes to infinity. Let $\{\tau_i\}$ be the break point fractions in the sense that

$$k_i = [\tau_i n], \quad 0 < \tau_1 < \ldots < \tau_R < 1. \tag{2.7}$$

Then, under suitable assumptions and linear series (2.1), Bai (1997), Proposition 9, showed that

$$n(\widehat{\tau}_i^{\,L} - \tau_i) = O_p(1), \tag{2.8}$$

where $\widehat{\tau}_i^{\,L} = \widehat{k}_i^L/n$. In the other direction, *small shifts* are considered in the sense that $\Delta_i = \Delta_i(n) = \nu_n \widetilde{\Delta}_i$, $i = 1, \ldots, R$, where

$$\nu_n \to 0, \text{ but } n^{1/2-\delta}\nu_n \to \infty \tag{2.9}$$

for some $\delta \in (0, 1/2)$ as $n \to \infty$. In this case, it is known (Bai (1997), Proposition 8) that

$$n\Delta_i^2(\widehat{\tau}_i^L - \tau_i) = O_p(1), \tag{2.10}$$

which implies consistency of the break point fractions with a slower convergence rate than in (2.8) where the shifts are fixed.

2.2. CUSUM method

The celebrated CUSUM statistic is defined as

$$\mathrm{CUSUM}(k) = \frac{1}{\sqrt{n}}\left(\sum_{j=1}^{k} X_j - \frac{k}{n}\sum_{j=1}^{n} X_j\right). \tag{2.11}$$

The respective break point estimator is

$$\widehat{k}^C = \underset{1\le k\le n}{\mathrm{argmax}} |\mathrm{CUSUM}(k)|. \tag{2.12}$$

It can be seen that

$$\mathrm{CUSUM}(k) = \frac{k}{n}\left(1 - \frac{k}{n}\right)\sqrt{n}\left(\overline{X}_k - \overline{X}_k^*\right), \tag{2.13}$$

where $\overline{X}_k$ and $\overline{X}_k^*$ are given in (2.4).

Denoting

$$\mathrm{adjCUSUM}(k) = \left(\frac{k}{n}\left(1 - \frac{k}{n}\right)\right)^{-1/2} \mathrm{CUSUM}(k), \tag{2.14}$$

note from (2.6) that the LSE estimator is

$$\widehat{k}^L = \underset{1\le k\le n-1}{\mathrm{argmax}} \left|\mathrm{adjCUSUM}(k)\right|. \tag{2.15}$$

Therefore, the CUSUM estimator $\widehat{k}^C$ differs from the LSE estimator $\widehat{k}^L$ only by a multiplicative factor $(k/n(1-k/n))^{-1/2}$. For this reason, the LSE method is sometimes also called the adjusted CUSUM. The two estimators $\widehat{k}^L$ and $\widehat{k}^C$ behave similarly, but it is also known (Csörgő and Horváth (1997)) that the LSE estimator performs better when the true break occurs near the boundaries.

The asymptotics of the CUSUM estimator are also well studied (see, for example, Csörgő and Horváth (1997)). However, most of the results are in the so-called AMOC (at most one change) framework. For example, in the case of small shifts (2.9) and under the CM-1 model, the CUSUM estimator satisfies

$$n\Delta_1^2(\widehat{\tau}_1^C - \tau_1) = O_p(1), \tag{2.16}$$

where $\hat{\tau}_1^C = \widehat{k}_1^C/n$. The limiting distribution of $n\Delta_1^2(\widehat{\tau}_1^C - \tau_1)$ is also known (Csörgő and Horváth (1997), Theorem 4.1.5). Asymptotics of the CUSUM estimator for fixed shifts is apparently not established yet.

3. Stopping rules

As referred to in the previous section, a common way to find multiple breaks is to repeat the procedure for finding breaks to subsamples of the series before and after an already estimated break point. This way of discovering multiple breaks needs a stopping rule. There are several ways to introduce such a rule.

One basic approach is to test within each subsample (starting with the whole sample) whether a break is present. If the hypothesis of no break is not rejected, the subsample is taken to have no break. Otherwise, a break is estimated and testing is repeated on the two subsamples (before and after the estimated break). This approach is illustrated in Section 3.1 through the sup-F test considered in Bai (1997). It is also sometimes referred to as a binary segmentation (Sen and Srivastava (1975)).

Another approach is to follow the binary segmentation, but also to introduce an order in which the breaks are found. This can be thought as a sequential testing procedure for additional breaks, stopping the first time the hypothesis of a certain number of breaks is not rejected. In this sequential way of finding breaks, a test for an additional break is usually based on a "global" statistic. This approach is illustrated through the BHKS stopping rule in Section 3.2.

In Section 3.3, we discuss our stopping rule based on the local Whittle estimation. It is introduced in the case where the multiple breaks are found sequentially. Finally, the tests discussed in Sections 3.1 and 3.2 involve estimation of the so-called long-run variance, which is discussed in Section 3.4.

Remark 3.1. *An alternative way to find multiple breaks is through a simultaneous procedure (as in Bai and Perron (1998)), for example in conjunction with a variable selection (as in Yao (1988), Liu, Wu and Zidek (1997), Lavielle (2005)). We work with a binary segmentation and sequential procedures because of their computational efficiency, and a relative ease of technical manipulation.*

3.1. Sup-F rule for a binary segmentation

Within each segment of a binary segmentation, the presence of an additional break could be tested by using a variety of tests. One such natural test is the sup-F test proposed by Bai (1997). If the sample on a segment of interest is denoted $\{X_1, \ldots, X_n\}$, the sup-F test statistic is defined as

$$\sup F_n = \sup_{n\eta \le k \le n(1-\eta)} \frac{SST_n - SSE_n(k)}{\widehat{\sigma}^2}, \tag{3.1}$$

where

$$SST_n = \sum_{j=1}^{n} (X_j - \overline{X}_n)^2,$$

$$SSE_n(k) = \sum_{j=1}^{k} (X_j - \overline{X}_k)^2 + \sum_{j=k+1}^{n} (X_j - \overline{X}_k^*)^2$$

are, respectively, the sums of squared errors under the hypothesis of no break and a single break at time k ($\overline{X}_k$ and $\overline{X}_k^*$ are defined in (2.4)). $\widehat{\sigma}^2$ in (3.1) is a consistent estimator of the long-run variance of $\{X_1, \ldots, X_n\}$ under no break, and is discussed in more detail in Section 3.4 below. $\eta \in (0, 1/2)$ is taken in (3.1) to avoid the divergence of the statistic at the boundaries.

Under suitable assumptions on a linear series $\{\epsilon_j\}$ in (2.1) and supposing no break, it is known (Bai (1997), Lemma 10) that

$$\sup F_n \xrightarrow{d} \sup_{\eta \le t \le (1-\eta)} \frac{|B^0(t)|^2}{t(1-t)}, \tag{3.2}$$

where $\{B^0(t)\}_{t\in[0,1]}$ is a standard Brownian bridge, and $\xrightarrow{d}$ denotes convergence in distribution. The sup-F test against an additional break is then based on the critical values of the limiting distribution in (3.2).

The binary segmentation procedure together with the sup-F test yield the number of estimated breaks $\widehat{R}$. The following is the consistency result.

Theorem 3.1 (Bai (1997), Proposition 11)**.** *Suppose the CM-R model (1.4), fixed shifts Δ_i and the breaks k_i satisfying (2.7). Let $\{\epsilon_j\}$ be a linear sequence in (2.1). Suppose the number of breaks $\widehat{R}$ is estimated using the sup-F test with the size of the test $\alpha = \alpha(n)$ satisfying $\alpha(n) \to 0$ but $\liminf_{n\to\infty} n\alpha(n) > 0$. Then*

$$P\left(\widehat{R} = R\right) \to 1, \qquad \text{as } n \to \infty. \tag{3.3}$$

Remark 3.2. *Even if Theorem 3.1 (Proposition 11, Bai (1997)) assumes fixed shifts, its proof can be adapted to cover the case of small shifts.*

3.2. BHKS rule for a sequential procedure

It may be convenient to introduce an order or a sequential procedure in which breaks are found in a binary selection. A basic idea for doing this is to compare the values of a test statistic across different segments and to test for an additional break within the segment with the largest statistic value. We illustrate this with the CUSUM statistic used in Berkes et al. (2006), though the sup-F (Bai (1997), p. 330) and other statistics could also be used. When the CUSUM statistic is used in a sequential procedure as in Berkes et al. (2006), we will refer to the corresponding stopping rule as the *BHKS rule*.

To explain the BHKS stopping rule, let

$$T(s\!:\!e) = \frac{1}{\widehat{\sigma}} \max_{s+1\le k\le e} |\mathrm{CUSUM}(k)| \tag{3.4}$$

be the largest value of the CUSUM statistic (2.11) normalized by a square root of a consistent estimator $\widehat{\sigma}^2$ of the long-run variance (Section 3.4 below), both based on a series $\{X_{s+1}, X_{s+2}, \ldots, X_e\}$. The BHKS stopping rule is obtained from a sequential testing of the hypothesis H_0: CM-R against H_1: CM-$(R+1)$ for increasing R based on the test statistic (3.4). More specifically, when testing H_0: CM-0 (no break) against H_1: CM-1, one expects under mild assumptions (Berkes et al. (2006), Theorem 3.1) that

$$T(0\!:\!n) \xrightarrow{d} \sup_{0\le t\le 1} |B^0(t)|, \tag{3.5}$$

under H_0, where $\{B^0(t)\}_{0\le t\le 1}$ is a standard Brownian bridge. If H_0: CM-0 is not rejected based on the asymptotic result (3.5), one sets $\widehat{R}=0$. Otherwise, let $\widehat{k}^C$ be the break obtained through (2.12) based on the series $\{X_1,\dots,X_n\}$ and proceed by testing H_0: CM-1 against H_1: CM-2. Under H_0: CM-1, one similarly expects that

$$\max\left\{T(0\!:\!\widehat{k}_1^C),T(\widehat{k}_1^C\!:\!n)\right\}\xrightarrow{d}\max_{i=1,2}\sup_{0\le t\le 1}|B_i^0(t)|, \tag{3.6}$$

where $\{B_i^0(t)\}_{0\le t\le 1}$, $i=1,2$, are two independent standard Brownian bridges. If H_0: CM-1 is not rejected based on the asymptotic result (3.6), one sets $\widehat{R}=1$. Otherwise, the next break point $\widehat{k}_2^C$ is found through (2.12) based on the series $\{X_1,\dots,X_{\widehat{k}_1^C}\}$ or $\{X_{\widehat{k}_1^C+1},\dots,X_n\}$ depending on $T(0:\widehat{k}_1^C)>T(\widehat{k}_1^C:n)$ or $T(\widehat{k}_1^C\!:\!n)>T(0\!:\!\widehat{k}_1^C)$, respectively. This sequential procedure is continued till the first R for which H_0: CM-R is not rejected against H_1: CM-$(R+1)$, and the value of R is taken for $\widehat{R}$.

If $\widehat{R}$ is thus obtained by using the BHKS rule, the corresponding consistency result has not been stated explicitly or proved rigorously (though meant implicitly in Berkes et al. (2006) if their tests are consistent and the size of the tests decreases). If $\widehat{R}$ is obtained through a sequential procedure using the sup-F test, the consistency follows from the results of Bai (1997) (as mentioned on p. 331).

Under LRD, it is expected that $\widehat{R}\to\infty$ (in probability). Berkes et al. (2006) proved the divergence of the test statistic for LRD series when testing for H_0: CM-1, and implicitly implied that the same happens when testing for H_0: CM-R. The latter fact implies that $\widehat{R}\to\infty$.

3.3. Local Whittle rule

We propose here a stopping rule in a sequential procedure based on the local Whittle estimation of the LRD (or SRD) parameter. The local Whittle estimator of the LRD (or SRD) parameter is defined as (Robinson (1995a))

$$\widehat{d}_{lw}=\underset{d\in[\Theta_1,\Theta_2]}{\operatorname{argmin}}\left\{\log\left(\frac{1}{m}\sum_{l=1}^m\omega_l^{2d}I(\omega_l)\right)-2d\frac{1}{m}\sum_{l=1}^m\log\omega_l\right\}, \tag{3.7}$$

where $-1/2<\Theta_1<0<\Theta_2<1/2$ are fixed, m denotes the number of lower frequencies used in estimation, and

$$I(\omega_l)=\frac{1}{2\pi n}\left|\sum_{j=1}^n Y_je^{-ij\omega_l}\right|^2 \tag{3.8}$$

is the periodogram at Fourier frequencies $\omega_l=2\pi l/n$.

The number of breaks $\widehat{R}$ with the *LW (Local Whittle) rule* is found through a sequential procedure as follows. Under the CM-0 model and suitable assumptions (Robinson (1995a)), one expects that

$$\sqrt{m}\widehat{d}_{lw} \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{4}\right). \tag{3.9}$$

Under the CM-1 model, one expects that $\sqrt{m}\widehat{d}_{lw} \xrightarrow{p} +\infty$. Then, use these asymptotic results to test for H_0: CM-0 against H_1: CM-1. If H_0: CM-0 is not rejected, set $\widehat{R} = 0$. If H_0 is rejected, find the first break $\widehat{k}_1^L$ through the LSE method in (2.6). Then, proceed to testing for H_0: CM-1 against H_1: CM-2 by using (3.9) but where $\widehat{d}_{lw}$ is now based on the residual series

$$R_j^{(1)} = X_j - \widehat{X}_j^{(1)}, \quad j = 1, \dots, n, \tag{3.10}$$

where $\widehat{X}_j^{(1)}$, $1 \leq j \leq \widehat{k}_1^L$, is a constant mean level till $\widehat{k}_1^L$ and $\widehat{X}_j^{(1)}$, $\widehat{k}_1^L + 1 \leq j \leq n$, is a constant mean level after $\widehat{k}_1^L$. If H_0: CM-1 is not rejected, set $\widehat{R} = 1$. Otherwise, find the second break $\widehat{k}_2^L$ through the LSE method (2.15) based on the series $\{X_1, \dots, X_{\widehat{k}_1^L}\}$ or $\{X_{\widehat{k}_1^L+1}, \dots, X_n\}$ depending on $S(0 : \widehat{k}_1^L) > S(\widehat{k}_1^L : n)$ or $S(\widehat{k}_1^L : n) > S(0 : \widehat{k}_1^L)$, respectively, where

$$S(s : e) = \max_{s+1 \leq k < e} |\mathrm{adjCUSUM}(k)| \tag{3.11}$$

is the largest adjusted CUSUM statistic (or equivalently, the smallest sum of squared errors) based on $\{X_{s+1}, X_{s+2}, \dots, X_e\}$. Then, test for H_0: CM-2 against H_1: CM-3 by using (3.9) but where $\widehat{d}_{lw}$ is based on the residual series

$$R_j^{(2)} = X_j - \widehat{X}_j^{(2)}, \quad j = 1, \dots, n, \tag{3.12}$$

where $\widehat{X}_j^{(2)}$ corresponds to the constant mean levels of the series determined by the two breaks $\widehat{k}_1^L$ and $\widehat{k}_2^L$. This procedure is now continued till the first R for which H_0: CM-R is not rejected against H_1: CM-$(R+1)$, each time removing R constant local mean levels from the series and testing whether the LRD (or SRD) parameter is in the SRD regime. The resulting values of R is taken for $\widehat{R}$.

The following result establishes the consistency of the estimated number of breaks $\widehat{R}$ under CM models. The proof and the assumptions can be found in Appendix A. In particular, we work in the setting of small shifts.

Theorem 3.2. *Let $\widehat{R}$ be the estimated number of breaks as defined above. Suppose the size of the test $\alpha = \alpha(n)$ is such that $\alpha(n) \to 0$ but $\liminf_{n\to\infty} n\alpha(n) > 0$. Then, for the CM-R model under the assumptions stated in Appendix A,*

$$P\left(\widehat{R} = R\right) \to 1, \quad \text{as } n \to \infty. \tag{3.13}$$

The following result shows that the estimated number of breaks $\widehat{R}$ converges to $+\infty$ under LRD models. The proof and the assumptions can be found in Appendix B.

Theorem 3.3. *Let $\widehat{R}$ be the estimated number of breaks as defined above. Suppose the size of the test $\alpha = \alpha(n)$ is such that $\alpha(n) \to 0$ but $\liminf_{n\to\infty} n\alpha(n) > 0$. Then, for a LRD model under the assumptions in Appendix B,*

$$\widehat{R} \xrightarrow{p} \infty, \qquad \text{as } n \to \infty. \tag{3.14}$$

3.4. Long-run variance estimation

The test statistics used in Sections 3.1 and 3.2 involve a consistent estimator $\widehat{\sigma}^2$ of the so-called long-run variance $\sigma^2 = \sum_{h=-\infty}^{\infty} \gamma(h)$. We briefly review here several choices for $\widehat{\sigma}^2$.

A popular choice is the heteroskedasticity and autocorrelation consistent (HAC) estimator (Andrews (1991)) defined by

$$\widehat{\sigma}^2_{HAC} = \sum_{h=-(n-1)}^{n-1} K_q(h)\hat{\gamma}(h), \tag{3.15}$$

where $\hat{\gamma}(k) = n^{-1}\sum_{j=1}^{n-k}(X_{j+k} - \overline{X}_n)(X_j - \overline{X}_n)$ are sample autocovariances, $K_q(h) = K(h/q)/q$ is a scaled kernel function, e.g. with the Bartlett kernel $K(x) = (1-|x|)1_{\{|x|\le 1\}}$, and q is a bandwidth.

Since $\sigma^2 = 2\pi f(0)$ with a spectral density f, the long-run variance is also estimated in the frequency domain through

$$\widehat{\sigma}^2 = \sum_{l=-[n/2]}^{[n/2]} K_q(l) I(\omega_l), \tag{3.16}$$

where $I(\omega_l)$ is the periodogram in (3.8). Furthermore, Robinson (2005) proposed the memory autocorrelation consistent (MAC) estimator

$$\widehat{\sigma}^2_{MAC} = p(\widehat{d})\frac{1}{q}\sum_{l=1}^{q} \omega_l^{2\widehat{d}} I(\omega_l), \tag{3.17}$$

where $p(d)$ is a constant given by

$$p(d) = \begin{cases} 2\frac{\Gamma(1-2d)\sin(\pi d)}{d(1+2d)}, & d \neq 0, \\ 2\pi, \quad d = 0, \end{cases} \tag{3.18}$$

and $\widehat{d}$ is, for example, the local Whittle estimator of the LRD parameter.

4. Finite sample properties

In this section, we present the finite sample performance of the discussed methods through Monte Carlo simulations. The SRD errors $\{\epsilon_j\}$ in the CM model follow a Gaussian AR(1) model with unit variance

$$\epsilon_j = \rho\epsilon_{j-1} + u_j, \quad u_j \sim \text{i.i.d. } \mathcal{N}(0, 1-\rho^2) \tag{4.1}$$

for different parameter values $\rho \in \{.1, .2, \dots, .9\}$. Gaussian FARIMA(0,$d$,0) time series with $d \in \{.1, .15, \dots, .45\}$ are considered for LRD models, scaled to have a unit variance. Specifically, the following tests are considered.

- (sup-F) The sup-F rule for a binary segmentation in Section 3.1 is used with the Bartlett long-run variance estimator in (3.15). For the bandwidth parameter q, we take the data dependent bandwidth of Andrews (1991), which minimizes the asymptotic truncated MSE. For the AR(1) errors, the bandwidth is given by

$$\hat{q}_B = 1.1447 \left(\frac{4n\hat{\rho}^2}{(1-\hat{\rho}^2)^2} \right)^{1/3}, \tag{4.2}$$

where $\hat{\rho}$ is an OLS estimator given by

$$\hat{\rho} = \sum_{j=2}^{n} X_j X_{j-1} \Big/ \sum_{j=2}^{n} X_{j-1}^2. \tag{4.3}$$

- (LW) Our proposed method described in Section 3.3 is considered with the following tuning parameters. For the LW stopping rule, the number of frequencies m in the LW estimation is to be selected. Asymptotic theory suggests $m = O\left(n^{4/5}\right)$ when the errors are Gaussian (Robinson (1995a)). Henry (2001) suggested a data dependent bandwidth by minimizing the MSE for the LW estimator. For the AR(1) series, the bandwidth is

$$\widehat{m}_H = \left(\frac{3}{4\pi} \right)^{4/5} \left| \frac{-\hat{\rho}}{1-\hat{\rho}^2} \right|^{-2/5} n^{4/5}, \tag{4.4}$$

where $\hat{\rho}$ is the OLS estimator in (4.3). Observe, however, that as $\hat{\rho} \to 0$, $\widehat{m}_H$ diverges. To avoid the divergence, the actual bandwidth used here is

$$\widehat{m} = \min \left\{ \widehat{m}_H, n^{4/5} \right\}. \tag{4.5}$$

Note also that the LW rule can be improved by removing the bias in the LW estimation. It is well known (see, for example, Andrews and Sun (2004)) that the LW estimator for the AR(1) series has the following bias correction

$$\sqrt{m}\left(\widehat{d}_{lw} - d_{bias} \right) \xrightarrow{d} \mathcal{N}(0, 1/4), \tag{4.6}$$

where

$$d_{bias} = \frac{2\pi^2}{9} \left(\frac{m^2}{n^2} \right) \frac{2\rho}{(1-\rho)^2}.$$

Therefore, for the LW stopping rule, we use the result (4.6) instead of (3.9), together with the estimated $\widehat{m}$ in (4.5) and $\widehat{\rho}$ in (4.3).

- (CUSUM) The BHKS rule for a sequential procedure discussed in Section 3.2 is applied with the Bartlett long-run variance estimator (3.15) and the bandwidth in (4.2).

- (CUSUM-MAC) The BHKS rule with the MAC long-run variance estimator (3.17) is used. For the MAC estimator, $\hat{m}$ in (4.5) was also used to estimate the LRD parameter. This is because the MAC estimator also achieves the MSE with the bandwidth of order $n^{4/5}$ (Abadir, Distaso and Giraitis (2007)).
- (CUSUM-JX) In addition to the tests introduced above, we include the modification of the CUSUM test suggested by Juhl and Xiao (2009). Juhl and Xiao (2009) consider the residuals from nonparametric regression,

$$\hat{\epsilon}_j = X_j - \frac{1}{h}\sum_{i=1}^{n} K\left(\frac{j-i}{h}\right) X_i. \tag{4.7}$$

  The Bartlett long-run variance estimator with the bandwidth (4.2) is calculated from the residual series (4.7). The Epanechnikov kernel K is used in (4.7) with the bandwidth $h = 2n^{4/5}$ by following Juhl and Xiao (2009).
- (CUSUM-RO) This modification was recently proposed by Robbins et al. (2011), and it improves the power of the CUSUM test for correlated observations. The idea of modification is to transform the original correlated observations to uncorrelated ones. This is done by performing the CUSUM test on the residuals obtained by fitting a parametric ARMA(p, q) model to the given data.

All R programs for Monte Carlo simulations used in the section are available at `http://web.skku.edu/~crbaek`.

4.1. Simulations for CM models

Three different models are considered here:

- (CM-0) $X_j = \epsilon_j$,
- (CM-1) $X_j = .5 1_{\{[n/2]<j\le n\}} + \epsilon_j$,
- (CM-2) $X_j = .5 1_{\{[n/3]<j\le n\}} - .5 1_{\{[2n/3]<j\le n\}} + \epsilon_j$,

where $\{\epsilon_j\}$ are the AR(1) errors given in (4.1). The sample size is chosen as $n = 2,000$ and all results are based on 1,000 replications. The CM-2 model will be considered only in Section 4.3 below.

First, in Table 1, we report the sizes of the tests for H_0: CM-0 with significance level $\alpha = .05$. When the correlation parameter ρ is small, all methods achieve

Table 1
Size of test for H_0: CM-0

ρ	.1	.2	.3	.4	.5	.6	.7	.8	.9
sup-F	0.043	0.055	0.050	0.061	0.059	0.044	0.038	0.033	0.024
CUSUM	0.036	0.056	0.054	0.065	0.065	0.058	0.050	0.054	0.054
CUSUM-JX	0.040	0.059	0.060	0.074	0.076	0.071	0.067	0.074	0.083
CUSUM-RO	0.034	0.050	0.049	0.059	0.056	0.046	0.047	0.043	0.045
CUSUM-MAC	0.053	0.070	0.072	0.099	0.104	0.101	0.118	0.138	0.205
LW	0.045	0.043	0.055	0.051	0.055	0.058	0.081	0.067	0.097

Table 2
Power of test for H_0: CM-0 against H_1: CM-1

ρ	.1	.2	.3	.4	.5	.6	.7	.8	.9
sup-F	1.000	1.000	1.000	1.000	1.000	0.996	0.947	0.790	0.336
CUSUM	1.000	1.000	1.000	1.000	1.000	0.999	0.985	0.926	0.595
CUSUM-JX	1.000	1.000	1.000	1.000	1.000	0.999	0.986	0.948	0.680
CUSUM-RO	1.000	1.000	1.000	1.000	1.000	1.000	0.984	0.927	0.578
CUSUM-MAC	1.000	1.000	1.000	1.000	1.000	1.000	0.990	0.967	0.775
LW	1.000	1.000	0.998	0.994	0.957	0.896	0.760	0.569	0.353

the nominal significance level. As ρ is approaching 1, the CUSUM and CUSUM-RO methods perform best while the sup-F test is too conservative (empirical size is 2.4%). Empirical sizes are moderately distorted for the CUSUM-JX and LW, 8.3% and 9.7%, respectively. The CUSUM-MAC test shows the largest size distortion.

Table 2 shows the power of the tests for H_0: CM-0 against H_1: CM-1. All tests considered have excellent power up to moderate correlations ρ. When the correlations are very strong, the power is still acceptable but observe that the CUSUM-based methods perform better than the sup-F and LW.

4.2. Simulations for LRD models

We consider the empirical power of the tests against LRD alternatives. When LRD time series are considered, all tests are expected to reject CM models. Here, we only report the empirical power of the tests at two stages where the null hypothesis is either CM-0 or CM-1.

First, the tests of H_0: CM-0 against H_1: LRD are considered in Table 3. Observe that the power of test is increasing as the LRD parameter d increases for all the tests considered here. On the other hand, when the LRD parameter d is small to moderate, the LW method outperforms both the sup-F and CUSUM-based methods.

At the second stage, the empirical power for H_0: CM-1 against H_1: LRD is reported in Table 4. (Note that the sup-F test is excluded because it is not sequential.) The overall power is smaller than that in Table 3. For the LW method, however, the power remains close to 1 for most of the LRD alternatives considered.

Table 3
Power of test for H_0: CM-0 against H_1: FARIMA(0, d, 0)

d	.1	.15	.2	.25	.3	.35	.4	.45
sup-F	0.296	0.429	0.544	0.633	0.687	0.735	0.773	0.772
CUSUM	0.302	0.401	0.536	0.632	0.656	0.735	0.781	0.781
CUSUM-JX	0.306	0.409	0.549	0.646	0.673	0.750	0.800	0.775
CUSUM-RO	0.348	0.521	0.696	0.806	0.881	0.930	0.963	0.973
CUSUM-MAC	0.429	0.638	0.806	0.873	0.948	0.941	0.872	0.680
LW	0.873	0.994	0.999	1.000	0.999	1.000	1.000	1.000

TABLE 4
Power of test for H_0: CM-1 against H_1: FARIMA(0, d, 0)

d	.1	.15	.2	.25	.3	.35	.4	.45
CUSUM	0.220	0.368	0.466	0.534	0.598	0.662	0.679	0.694
CUSUM-JX	0.235	0.397	0.486	0.563	0.629	0.705	0.705	0.715
CUSUM-RO	0.263	0.488	0.630	0.747	0.846	0.891	0.929	0.950
CUSUM-MAC	0.392	0.620	0.777	0.864	0.915	0.926	0.886	0.798
LW	0.779	0.962	0.997	0.998	0.998	1.000	1.000	1.000

4.3. Estimated number of breaks for distinguishing CM and LRD models

Table 5 presents empirical frequencies of the estimated numbers of breaks for the CM-2 model. Observe that when the correlation in the errors is moderate, the estimated numbers of breaks are highly centered around the true number of breaks $R = 2$ for all methods. In the case of larger correlations, on the other

TABLE 5
Empirical frequencies (in percentage) of the estimated numbers of breaks for CM-2

	sup-F						CUSUM					
$\rho\backslash\widehat{R}$	0	1	2	3	4	≥ 5	0	1	2	3	4	≥ 5
.1	0.0	0.0	78.0	18.7	2.6	0.7	0.0	0.0	94.3	5.7	0.0	0.0
.2	0.0	0.4	77.5	18.5	3.2	0.4	0.1	0.0	94.7	5.0	0.2	0.0
.3	0.0	4.1	76.7	16.7	2.3	0.2	0.4	0.0	95.1	4.4	0.0	0.1
.4	0.0	15.0	69.6	13.3	2.1	0.0	3.1	0.1	92.5	4.3	0.0	0.0
.5	0.0	36.2	52.1	10.4	1.1	0.2	13.2	0.4	82.2	4.1	0.1	0.0
.6	0.0	58.9	35.4	5.1	0.6	0.0	31.2	1.9	65.0	1.9	0.0	0.0
.7	0.0	77.8	19.8	2.3	0.1	0.0	49.6	7.1	42.3	1.0	0.0	0.0
.8	0.0	95.3	4.4	0.3	0.0	0.0	74.4	10.8	14.7	0.1	0.0	0.0
.9	0.0	99.7	0.3	0.0	0.0	0.0	89.2	7.9	2.9	0.0	0.0	0.0
	CUSUM-JX						CUSUM-RO					
$\rho\backslash\widehat{R}$	0	1	2	3	4	≥ 5	0	1	2	3	4	≥ 5
.1	0.0	0.0	94.2	5.8	0.0	0.0	0.0	0.0	95.4	4.6	0.0	0.0
.2	0.1	0.0	94.5	5.3	0.1	0.0	0.1	0.0	96.6	3.3	0.0	0.0
.3	0.1	0.0	94.9	4.9	0.0	0.1	0.1	0.0	96.2	3.6	0.0	0.1
.4	1.8	0.1	93.6	4.4	0.1	0.0	1.9	0.0	95.1	3.0	0.0	0.0
.5	7.1	0.4	87.9	4.5	0.1	0.0	7.6	0.6	88.8	3.0	0.0	0.0
.6	19.7	2.0	75.4	2.9	0.0	0.0	22.9	2.4	72.7	1.9	0.1	0.0
.7	36.2	6.9	55.4	1.5	0.0	0.0	40.9	6.6	51.1	1.3	0.1	0.0
.8	61.5	11.4	26.6	0.5	0.0	0.0	66.1	11.3	22.3	0.3	0.0	0.0
.9	79.8	11.2	9.0	0.0	0.0	0.0	86.4	9.3	4.3	0.0	0.0	0.0
	CUSUM-MAC						LW					
$\rho\backslash\widehat{R}$	0	1	2	3	4	≥ 5	0	1	2	3	4	≥ 5
.1	0.0	0.0	91.7	7.8	0.5	0.0	0.0	1.0	95.8	2.6	0.4	0.2
.2	0.0	0.0	89.4	9.3	1.2	0.1	0.1	3.1	91.7	4.2	0.4	0.5
.3	0.0	0.0	89.9	9.4	0.4	0.3	0.6	7.0	89.2	2.5	0.7	0.0
.4	0.2	0.0	89.1	9.8	0.9	0.0	1.6	14.1	78.9	4.4	0.5	0.5
.5	1.4	0.4	84.3	12.0	1.7	0.2	5.7	24.5	65.7	3.1	0.7	0.3
.6	4.7	0.6	79.8	12.3	2.5	0.1	15.7	31.0	48.9	3.5	0.4	0.5
.7	13.1	5.1	67.4	11.8	2.3	0.3	31.4	34.6	31.2	2.4	0.3	0.1
.8	27.8	8.9	48.8	11.6	2.6	0.3	49.7	28.9	18.2	2.5	0.4	0.3
.9	46.7	13.3	29.2	9.4	1.3	0.1	67.9	18.4	10.4	2.3	0.6	0.4

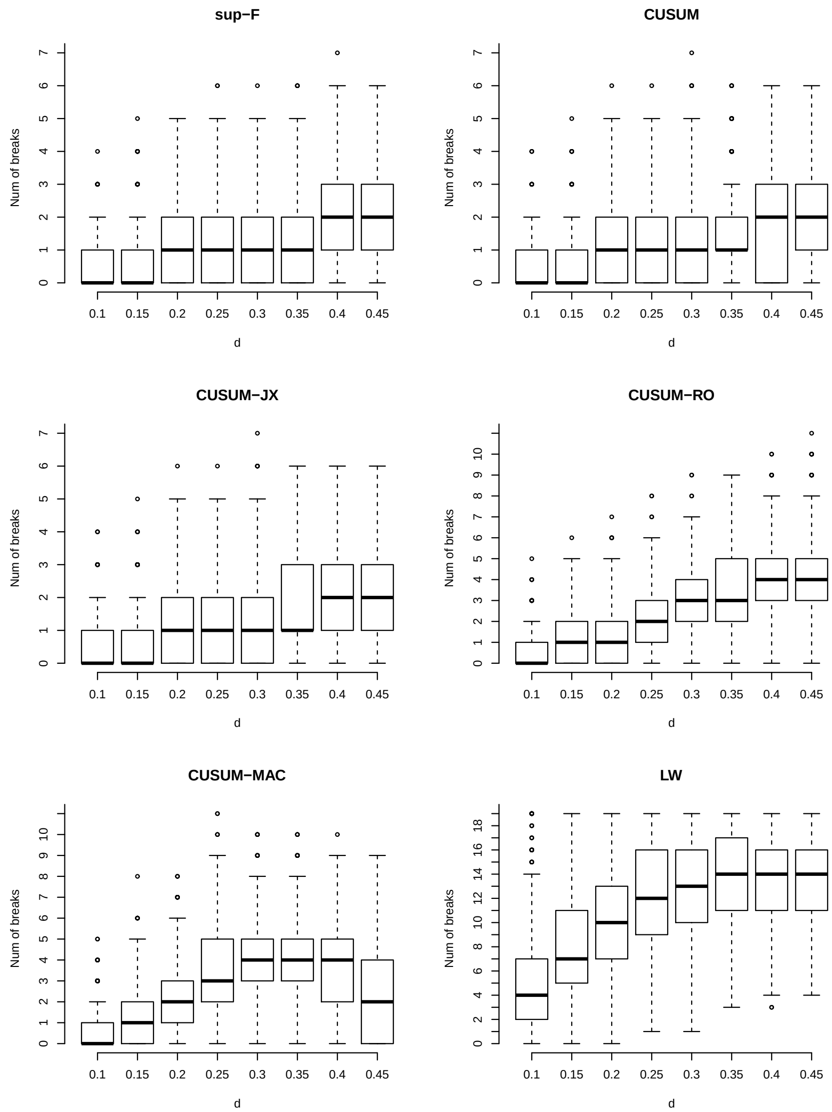

FIG 1. *Estimated number of breaks for FARIMA(0,d,0).*

hand, all methods tend to underestimate the true number of breaks, but the results are quite similar amongst all the methods considered. The results for the CM-0 and CM-1 models (not reported here) are similar to or even better than those for the CM-2 model.

The numbers of estimated breaks for LRD series are reported in Figure 1. There is now a big difference in the results between the LW method and the available methods based on the sup-F or CUSUM tests. (Note a different scale on the vertical axis for the LW method.) For all the LRD parameter values d considered, the LW method finds a much larger number of estimated breaks. We note again that this is due to the LW method having a better power against LRD.

The above findings suggest a way to distinguish CM and LRD models. For a given observed time series, apply all the procedures including the LW method discussed in Section 3.3, and compare the estimated numbers of breaks. If the estimated numbers of breaks are not that different across all the procedures, the observed time series is consistent with a CM model. On the contrary, if the LW method finds a much larger number of estimated breaks than the sup-F or CUSUM-based methods, the observed time series is consistent with LRD. This procedure is formalized and applied to several real time series in the next sections.

4.4. Bootstrap procedure to test CM models against LRD

In Section 4.3, we argued that CM and LRD models can be distinguished by comparing the numbers of estimated breaks. We formalize here this idea through a test, and use bootstrap to compute p-values. Using bootstrap in CM models has received attention only recently. See, for example, Hušková and Kirch (2008), Seijo and Sen (2011), and Chang and Perron (2014). However, to the best of our knowledge, bootstrapping the number of changes has not been considered yet. Justifying our use of bootstrap goes beyond the scope of this work and will be addressed in the future.

For the sake of simplicity, we will compare only the numbers of breaks estimated through the CUSUM and LW methods. Other CUSUM based methods can be considered similarly. Suppose that the number $\widehat{R}$ of breaks is estimated using the CUSUM method. Write the residuals as

$$\widehat{\epsilon}_j = X_j - \widehat{X}_j = X_j - \left(\widehat{\mu} + \sum_{r=1}^{\widehat{R}} \widehat{\Delta}_r 1_{\{\widehat{k}_r < j \le n\}}\right), \quad j = 1, \dots, n. \tag{4.8}$$

Further, denote the centered residuals as

$$\widehat{\epsilon}_j^* = \widehat{\epsilon}_j - n^{-1} \sum_{i=1}^{n} \widehat{\epsilon}_i.$$

We now use the block bootstrap of block size L to obtain bootstrap residuals (see, for example, Kunsch (1989), Lahiri (1999)). The block length L is taken as

$$L = 2 \min\{h \ge 1 : |\widehat{\gamma}_{\hat{\epsilon}}(h)| \le 1.96/\sqrt{n}\},$$

where $\widehat{\gamma}_{\hat{\epsilon}}(h)$ is the sample ACF function of the residual series $\{\widehat{\epsilon}_j, j = 1, \dots, n\}$. By superimposing the estimated CM model and the bootstrap residuals, we obtain a bootstrap sample of observations $\{X_1^*, \dots, X_n^*\}$. Reapplying the CUSUM and LW methods to the bootstrap sample $\{X_1^*, \dots, X_n^*\}$ gives the estimated

Table 6
Empirical size of test based on bootstrapping for CM-2.

ρ	.1	.2	.3	.4	.5	.6	.7	.8	.9
Bootstrap	0.006	0.020	0.012	0.014	0.016	0.010	0.012	0.010	0.022

Table 7
Empirical power of test based on bootstrapping for FARIMA$(0,d,0)$.

d	.1	.15	.2	.25	.3	.35	.4	.45
Bootstrap	0.084	0.266	0.438	0.692	0.896	0.948	0.978	0.982

numbers of breaks $\widehat{R}_C$ and $\widehat{R}_W$, respectively. By repeating n_B times, we produce bootstrap estimated numbers of breaks denoted by $\{\widehat{R}_C^{(1)},\ldots,\widehat{R}_C^{(n_B)}\}$ and $\{\widehat{R}_W^{(1)},\ldots,\widehat{R}_W^{(n_B)}\}$, respectively, as well as bootstrap differences in the estimated numbers $\{\widehat{R}_W^{(1)}-\widehat{R}_C^{(1)},\ldots,\widehat{R}_W^{(n_B)}-\widehat{R}_C^{(n_B)}\}$. We will reject the null hypotheses of H_0: CM model in favor of LRD for a given significance level α if the $100\alpha\%$ sample quantile of the differences $\widehat{R}_W^{(i)}-\widehat{R}_C^{(i)}$, $i=1,\ldots,n_B$, is greater than zero. This is equivalent to rejecting the null hypotheses if

$$\frac{1}{n_B}\sum_{i=1}^{n_B} 1_{\{\widehat{R}_W^{(i)}-\widehat{R}_C^{(i)}\le 0\}} < \alpha.$$

We will further apply a bias correction by centering the difference, with the final decision rule to reject the null hypotheses in favor of LRD if

$$\frac{1}{n_B}\sum_{i=1}^{n_B} 1_{\{\widehat{R}_W^{(i)}-\widehat{R}_C^{(i)}-\min(0,\overline{R}_W-\overline{R}_C)\ \le 0\}} < \alpha, \tag{4.9}$$

where $\overline{R}_W=n_B^{-1}\sum_{i=1}^{n_B}\widehat{R}_W^{(i)}$ and $\overline{R}_C=n_B^{-1}\sum_{i=1}^{n_B}\widehat{R}_C^{(i)}$. The bias correction is to account for possibly different means of the distributions of the number of breaks obtained by the LW and CUSUM methods under CM models. Indeed, note from Table 5, for example, that there is a slightly larger proportion of breaks less than 2 when using the LW method.

Table 6 presents the empirical size of the bootstrap test for the CM-2 model with the Gaussian AR(1) errors described in Section 4.1. The results are based on the sample size of $n=2,000$ with $n_B=1,000$ and 500 replications under 5% significance level. The empirical power is examined by generating FARIMA$(0,d,0)$ series of length $n=2,000$. Observe that the empirical sizes are less than the 5% nominal level in all cases considered, indicating that the bootstrapping procedure is conservative. However, the empirical power approaches 1 as d increases. For example, when $d=.35$, the empirical power is about 95%.

5. Applications to real data

We first consider several time series of volatility of stock indices. As in Stărică and Granger (2005) (and similarly to McCloskey and Perron (2013) and others),

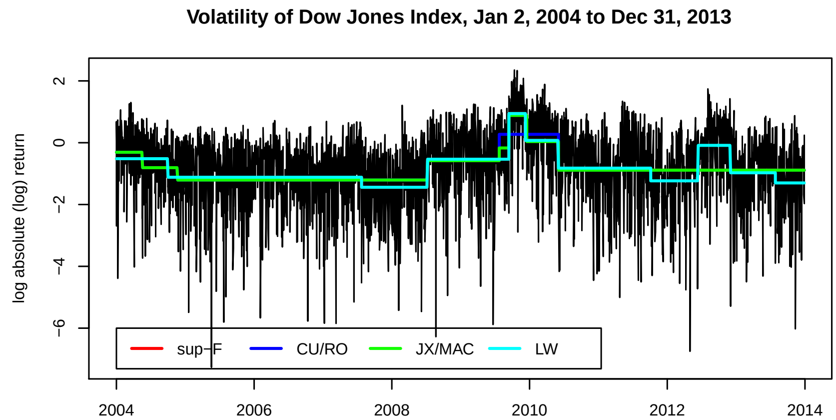

Fig 2. *Estimated CM models for DJI volatility. CUSUM and CUSUM-JX find 5 breaks, sup-F finds 6 breaks, CUSUM-RO and CUSUM-MAC find 7 breaks. LW method, on the other hand, finds 10 breaks.*

we consider the logarithm of the absolute daily log-returns

$$\log |r_t| = \log |100(\log(P_t) - \log(P_{t-1}))| \, , \tag{5.1}$$

where P_t is the daily closing stock index. Two such indices are examined: S&P500 and Dow Jones Industrial (DJI) Average, both from the period of Jan 2, 2004 to Dec 31, 2013 (2515 observations).

Figure 2 shows a time series plot of the log absolute daily log-returns for DJI with the estimated CM models. The number of estimated breaks are slightly different amongst the CUSUM-based methods. For example, CUSUM and CUSUM-JX find 5 breaks, sup-F finds 6 breaks, CUSUM-RO and CUSUM-MAC find 7 breaks. The LW method, on the other hands, finds 10 breaks. Observe that the estimated CM models are similar amongst all the methods considered, except that the LW method finds more breaks in 2009–2012, which is roughly the period of the financial crisis. Our proposed bootstrap test based on the CUSUM and LW methods gives the p-value of .690, suggesting a CM model. The same conclusion was reached by other bootstrap tests, for example, the CUSUM-MAC and LW method gives the p-value of .589. The results for S&P500 are presented in Figure 3. All CUSUM-based and sup-F methods find 6 breaks while the LW method finds 12 breaks. But again, the estimated CM models are not that different, and the LW method finds more breaks in the period of the financial crisis 2008-2012. The bootstrap test based on CUSUM and LW gives the p-value of .682, hence suggesting a CM model. The same conclusion was reached by McCloskey and Perron (2013) for S&P 500 in the period from July 13, 1962 to March 25, 2004 and DJI in the period from March 4, 1957 to Oct 30, 2002, by using a different approach.

The next celebrated data set consists of the water levels of the Nile river measured at the Roda gauge during the years 622 to 1284 AD (663 observations).

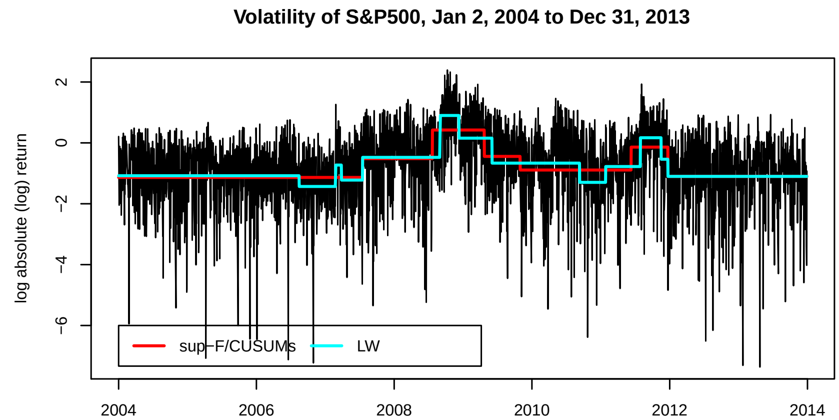

FIG 3. *Estimated CM models for S&P500 volatility. CUSUM-based and sup-F methods find 6 breaks while LW method finds 12 breaks.*

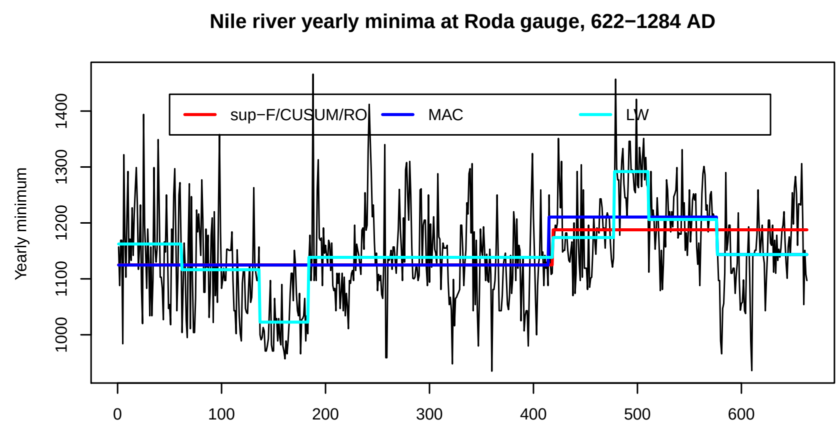

FIG 4. *Estimated CM models for the water levels of the Nile river. CUSUM-JX finds 0 breaks, sup-F, CUSUM and CUSUM-RO find 1 and CUSUM-MAC find 2 breaks. LW method, on the other hand, finds 7 breaks.*

For this time series, our proposed LW method finds 7 breaks while other methods find only 0 to at most 2 breaks (sup-F, CUSUM, CUSUM-RO find 1, CUSUM-JX finds 0 and CUSUM-MAC finds 2 breaks). The big (relative) difference in the estimated numbers of breaks suggests that the series is consistent with LRD. The bootstrap test based on CUSUM-MAC and LW gives the p-value of .078, and CUSUM and LW gives the p-value of .059. McCloskey and Perron (2013) also advocated LRD.

The next data set considers the quarterly congressional job approval in the US from 1969 to 2009. The total number of observations is 161. The congressional job approval is modeled as a LRD series in, for example, Byers, Davidson and

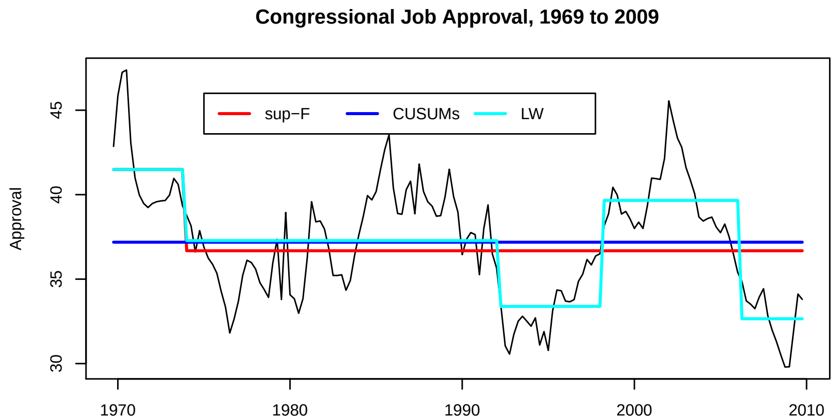

FIG 5. *Estimated CM models for the congressional approval in the US. CUSUM-based methods find 0 breaks, sup-F finds 1. LW method, on the other hand, finds 4 breaks.*

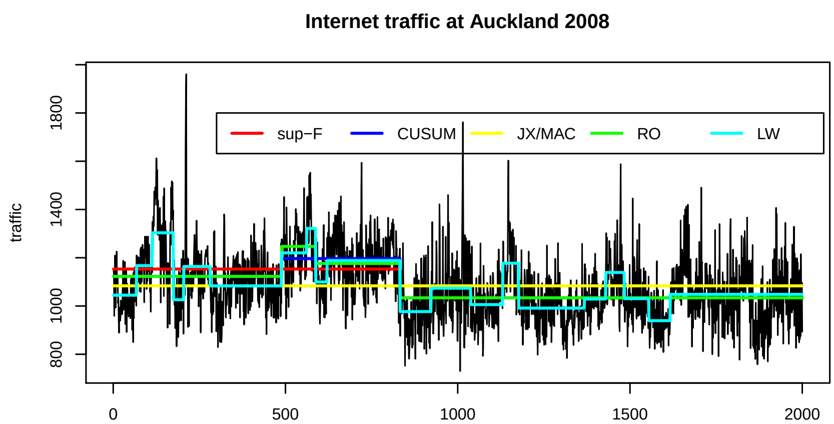

FIG 6. *Estimated CM models for Internet packet traffic at Auckland. CUSUM-JX finds 0 breaks, sup-F finds 1 break, CUSUM and CUSUM-MAC find 2 breaks, CUSUM-RO finds 3 breaks. For LW method, CM-19 is selected.*

Peel (1997), Box-Steffensmeier and Tomlinson (2000) and Lebo, Walker and Clarke (2000). Figure 5 presents the estimated CM models. The CUSUM-based method finds no break and the sup-F method finds only 1 break. Our proposed method finds four breaks and captures more rapid changes, suggesting LRD. However, the bootstrap test based on the sup-F and LW methods gives the p-value of .216. Hence, there is not enough evidence to reject CM models in favor of LRD. It is also plausible that the sample size is just not large enough to reject the null.

The last data set considered is the Internet packet traffic data collected from the popular Auckland IX depository for March 28, 2008 (the data can be downloaded at http://wand.net.nz/wits/auck/9/). Internet packet traffic is aggregated to 100 milliseconds. The difference between the LW method and the CUSUM-based methods is now striking. CUSUM-JX fails to find any break, sup-F finds 1 break, CUSUM and CUSUM-MAC find 2 breaks and CUSUM-RO finds 3 breaks. However, our proposed LW method finds 19 breaks. This suggests strongly that the series is more consistent with LRD. The bootstrap test gives the p-value of .001 for CUSUM and LW, and the p-value of less than .001 when CUSUM-MAC and LW methods are used, thus a strong evidence in favor of a LRD model.

6. Conclusions

In this work, we have introduced a procedure at distinguishing CM and LRD models. Our proposed method is based on the LW estimation of the LRD parameter from the residual series obtained by sequentially removing changes in mean. It leads to a consistent new stopping rule to estimate the number of breaks in CM models. A simulation study shows that our LW test has a better power against LRD than the available CUSUM and sup-F methods. This suggests that CM and LRD models can be distinguished by comparing the numbers of estimated breaks when using the LW method and the available sup-F or CUSUM methods. Our approach is illustrated on several real data series.

Appendix A: Proof of Theorem 3.2

We will assume the following conditions throughout this section.

Assumptions

(A1) Suppose the CM-R model (1.4) with a linear sequence $\{\epsilon_j\}$ satisfying (2.1) and also the assumptions of Theorem 2 of Robinson (1995a) with $H_0 = 1/2$.

(A2) $k_i = [n\tau_i]$ for some $\tau_i \in (0,1)$ such that $\tau_1 < \tau_2 < \ldots < \tau_R$.

(A3) $\Delta_i = \Delta_i(n) = \nu_n \widetilde{\Delta}_i$, $i = 1, \ldots, R$, where $\nu_n \to 0$, but $n^{1/2-\delta}\nu_n \to \infty$ for some $\delta \in (0, 1/2)$.

(A4) Suppose that the break points $\widehat{k}_i$ are estimated sequentially using the LSE method (Section 2), and the stopping rule based on (3.11). In particular,

$$\Delta_i^2(\widehat{k}_i - k_i) = O_p(1), \text{ or equivalently } n\Delta_i^2(\widehat{\tau}_i - \tau_i) = O_p(1)$$

(see (2.10)).

(A5) For the number of low frequencies m used in the LW estimation, $\log n/m \to 0$ as $n \to \infty$, and $m^2/(n\Delta_i^2) \to 0$ for all $i = 1, \ldots, R$.

(A6) There are no ties (in the asymptotic sense) when finding break points at each stage. For the first break, this means that

$$\lim(P)\nu_n^{-2}\left(U_n(\tau_i) - U_n(\tau_j)\right) < 0$$

for some i and all $i \neq j$, where $U_n(\tau) = n^{-1}SSE_n([n\tau])$ (see Bai (1997)).

Proof of Theorem 3.2. First, we will consider the case when $\widehat{R} < R$, and show that $P(\widehat{R} < R) \to 0$ as $n \to \infty$. Without loss of generality assume that $\widehat{R} = R-1$ since the other cases can be argued similarly only with additional terms. Since $\widehat{R} < R$, in view of Proposition 8 of Bai (1997), the $\widehat{R}$ estimated break point fractions are consistent estimators of $(R-1)$ true break point fractions. This means that there is a break point, say k_i, which failed to be detected in a sequential procedure. Denote the $(R-1)$ estimated break points as

$$\widehat{k}_1 < \widehat{k}_2 < \ldots < \widehat{k}_{i-1} < \widehat{k}_{i+1} < \ldots < \widehat{k}_R \tag{A.1}$$

and let $\mathcal{P} = \{0, 1, \ldots, i-1, i+1, \ldots, R+1\}$ be the index set. Then, the periodogram based on $R_j^{(R-1)} = X_j - \widehat{X}_j^{(R-1)}$ becomes

$$\begin{aligned} I_R(\omega_l) &= \left| \frac{1}{\sqrt{2\pi n}} \sum_{j=1}^{n} R_j^{(R-1)} e^{-ij\omega_l} \right|^2 \\ &= \left| \frac{1}{\sqrt{2\pi n}} \sum_{j=1}^{n} \epsilon_j e^{-ij\omega_l} + \frac{1}{\sqrt{2\pi n}} \sum_{j=\widehat{k}_{i-1}+1}^{\widehat{k}_{i+1}} (m_j - \widehat{X}_j^{(R-1)}) e^{-ij\omega_l} \right. \\ &\quad \left. + \frac{1}{\sqrt{2\pi n}} \sum_{p \in \mathcal{P}\setminus\{i-1\}} \sum_{j=\widehat{k}_p+1}^{\widehat{k}_{p+1}} (m_j - \widehat{X}_j^{(R-1)}) e^{-ij\omega_l} \right|^2, \end{aligned} \tag{A.2}$$

where $m_j = \mu + \sum_{r=1}^{R} \Delta_r 1_{\{k_r < j \le n\}}$ with the convention that $\widehat{k}_0 = 0$ and $\widehat{k}_{R+1} = n$. We need to understand the behavior of the local Whittle estimator $\widehat{d}_{lw}$ based on the periodogram $I_R(\omega_l)$. We will show that $\widehat{d}_{lw} \xrightarrow{p} \Theta_1$. As will be seen below, this yields the desired convergence $P(\widehat{R} < R) \to 0$.

Consider the regime $(\widehat{k}_{i-1}, \widehat{k}_{i+1}]$ where the true break point k_i is contained. Here, for the shortness sake, we will only consider the case when $\widehat{k}_{i-1} > k_{i-1}$, $\widehat{k}_{i+1} < k_{i+1}$. The other cases, for example when $\widehat{k}_{i-1} \le k_{i-1}$ and $\widehat{k}_{i+1} \ge k_{i+1}$, can be proved by using similar arguments. Observe that, in the case considered,

$$m_j - \widehat{X}_j^{(R-1)} = \begin{cases} -\Delta_i \dfrac{\widehat{k}_{i+1} - k_i}{\widehat{k}_{i+1} - \widehat{k}_{i-1}} - \dfrac{1}{\widehat{k}_{i+1} - \widehat{k}_{i-1}} \displaystyle\sum_{j=\widehat{k}_{i-1}+1}^{\widehat{k}_{i+1}} \epsilon_j, & \widehat{k}_{i-1} < j \le k_i, \\ \Delta_i \dfrac{k_i - \widehat{k}_{i-1}}{\widehat{k}_{i+1} - \widehat{k}_{i-1}} - \dfrac{1}{\widehat{k}_{i+1} - \widehat{k}_{i-1}} \displaystyle\sum_{j=\widehat{k}_{i-1}+1}^{\widehat{k}_{i+1}} \epsilon_j, & k_i < j \le \widehat{k}_{i+1}. \end{cases}$$

Therefore, the periodogram (A.2) can be written as

$$I_R(\omega_l) = \left| \frac{1}{\sqrt{2\pi n}} \sum_{j=1}^{n} \epsilon_j e^{-ij\omega_l} - \frac{\widehat{k}_{i+1} - k_i}{\widehat{k}_{i+1} - \widehat{k}_{i-1}} \frac{\Delta_i}{\sqrt{2\pi n}} \sum_{j=\widehat{k}_{i-1}+1}^{k_i} e^{-ij\omega_l} \right.$$

$$
\begin{aligned}
&+\frac{k_i-\widehat{k}_{i-1}}{\widehat{k}_{i+1}-\widehat{k}_{i-1}}\frac{\Delta_i}{\sqrt{2\pi n}}\sum_{j=k_i+1}^{\widehat{k}_{i+1}}e^{-ij\omega_l}\\
&-\frac{1}{\sqrt{2\pi n}}\sum_{j=\widehat{k}_{i-1}+1}^{\widehat{k}_{i+1}}\epsilon_j\frac{1}{\widehat{k}_{i+1}-\widehat{k}_{i-1}}\sum_{j=\widehat{k}_{i-1}+1}^{\widehat{k}_{i+1}}e^{-ij\omega_l}\\
&+\frac{1}{\sqrt{2\pi n}}\sum_{p\in\mathcal{P}\setminus\{i-1\}}\sum_{j=\widehat{k}_p+1}^{\widehat{k}_{p+1}}(m_j-\widehat{X}_j^{(R-1)})e^{-ij\omega_l}\Bigg|^2\\
=:&\;|x_1+x_2+x_3+x_4+x_5|^2
\end{aligned}
$$

and hence as

$$
I_R(\omega_l)=\frac{n\Delta_i^2}{2\pi}\left(\left|\frac{1}{n}\frac{\widehat{k}_{i+1}-k_i}{\widehat{k}_{i+1}-\widehat{k}_{i-1}}\sum_{j=\widehat{k}_{i-1}+1}^{k_i}e^{-ij\omega_l}\right|^2+\left|\frac{1}{n}\frac{k_i-\widehat{k}_{i-1}}{\widehat{k}_{i+1}-\widehat{k}_{i-1}}\sum_{j=k_i+1}^{\widehat{k}_{i+1}}e^{-ij\omega_l}\right|^2+\nu_l\right),\tag{A.3}
$$

where ν_l denotes the remaining term expressed as

$$
\frac{n\Delta_i^2}{2\pi}\nu_l=|x_1|^2+|x_4|^2+|x_5|^2+2\Re\left(\sum_{1\le j<k\le 5}x_j\bar{x}_k\right).
$$

It follows that

$$
\begin{aligned}
\widehat{d}_{lw}&=\operatorname*{argmin}_{\Theta_1\le d\le\Theta_2}\left\{\log\frac{1}{m}\sum_{l=1}^m\omega_l^{2d}I_R(\omega_l)-2d\frac{1}{m}\sum_{l=1}^m\log\omega_l\right\}\\
&=\operatorname*{argmin}_{\Theta_1\le d\le\Theta_2}\left\{\log\frac{1}{m}\sum_{l=1}^m l^{2d}I_R(\omega_l)-2d\frac{1}{m}\sum_{l=1}^m\log l\right\}\\
&=\operatorname*{argmin}_{\Theta_1\le d\le\Theta_2}\Bigg\{\log\sum_{l=1}^m l^{2d}\Bigg(\left|\frac{1}{n}\frac{\widehat{k}_{i+1}-k_i}{\widehat{k}_{i+1}-\widehat{k}_{i-1}}\sum_{j=\widehat{k}_{i-1}+1}^{k_i}e^{-ij\omega_l}\right|^2\\
&\quad+\left|\frac{1}{n}\frac{k_i-\widehat{k}_{i-1}}{\widehat{k}_{i+1}-\widehat{k}_{i-1}}\sum_{j=k_i+1}^{\widehat{k}_{i+1}}e^{-ij\omega_l}\right|^2+\nu_l\Bigg)-2d\frac{1}{m}\sum_{l=1}^m\log l\Bigg\}.
\end{aligned}\tag{A.4}
$$

We will first show that for some constants $c_1(l)^2$ and $c_2(l)^2$ specified below,

$$
\begin{aligned}
&\sup_{\Theta_1\le d\le\Theta_2}\Bigg|\sum_{l=1}^m l^{2d}\Bigg(\left|\frac{1}{n}\frac{\widehat{k}_{i+1}-k_i}{\widehat{k}_{i+1}-\widehat{k}_{i-1}}\sum_{j=\widehat{k}_{i-1}+1}^{k_i}e^{-ij\omega_l}\right|^2\\
&\quad+\left|\frac{1}{n}\frac{k_i-\widehat{k}_{i-1}}{\widehat{k}_{i+1}-\widehat{k}_{i-1}}\sum_{j=k_i+1}^{\widehat{k}_{i+1}}e^{-ij\omega_l}\right|^2+\nu_l\Bigg)-\sum_{l=1}^\infty l^{2d-2}\big(c_1(l)^2+c_2(l)^2\big)\Bigg|\xrightarrow{p}0.
\end{aligned}\tag{A.5}
$$

Observe that, by the assumptions (A3) and (A4) (and assuming for simplicity that $\tau_i n$ is an integer),

$$\begin{aligned}\sqrt{\frac{2\pi}{n\Delta_i^2}}|x_2| &= \left|\frac{1}{n}\frac{\widehat{k}_{i+1}-k_i}{\widehat{k}_{i+1}-\widehat{k}_{i-1}}\sum_{j=\widehat{k}_{i-1}+1}^{k_i} e^{-ij\omega_l}\right| \\ &= \frac{1}{n}\left|\frac{\widehat{\tau}_{i+1}-\tau_i}{\widehat{\tau}_{i+1}-\widehat{\tau}_{i-1}}\right|\frac{|e^{-i(\widehat{k}_{i-1}+1)\omega_l}-e^{-i(k_i+1)\omega_l}|}{|1-e^{-i\omega_l}|} \\ &= \left|\frac{\tau_{i+1}-\tau_i}{\tau_{i+1}-\tau_{i-1}}\right|\frac{|e^{-i\tau_{i-1}2\pi l}-e^{-i\tau_i 2\pi l}|}{2\pi l}+o_p(1) =: \frac{c_1(l)}{l}+o_p(1)\end{aligned}$$

uniformly in $l=1,\dots,m$. Similarly,

$$\begin{aligned}\sqrt{\frac{2\pi}{n\Delta_i^2}}|x_3| &= \left|\frac{1}{n}\frac{k_i-\widehat{k}_{i-1}}{\widehat{k}_{i+1}-\widehat{k}_{i-1}}\sum_{j=k_i+1}^{\widehat{k}_{i+1}} e^{-ij\omega_l}\right| \\ &= \left|\frac{\tau_i-\tau_{i-1}}{\tau_{i+1}-\tau_{i-1}}\right|\frac{|e^{-i\tau_i 2\pi l}-e^{-i\tau_{i+1}2\pi l}|}{2\pi l}+o_p(1) =: \frac{c_2(l)}{l}+o_p(1)\end{aligned}$$

uniformly in $l=1,\dots,m$. For the remainder term ν_l, note first that

$$x_1 = O_p(1) \tag{A.6}$$

uniformly in $l=1,2,\dots,m$, since $\sup_{1\le l\le m} E|x_1|<\infty$ by Theorem 2 in Robinson (1995b); see also (3.16) in Robinson (1995a). It follows from the above that

$$x_2 = O_p\left(\sqrt{\frac{n\Delta_i^2}{l^2}}\right), \quad x_3 = O_p\left(\sqrt{\frac{n\Delta_i^2}{l^2}}\right).$$

Note also that

$$x_4 = O_p\left(\frac{1}{l}\right)$$

from the assumption (A1) by using a standard martingale central limit theorem for $\{\epsilon_j\}$ (see Hall and Heyde (1980), Theorem 5.5, pp. 141-146, and note also that $\sum_{i=0}^{\infty} i|\psi_i|<\infty$ implies (5.37) of Hall and Heyde (1980)).

For x_5, consider the regime $(\widehat{k}_p,\widehat{k}_{p+1}]$. Observe that, for example, if $\widehat{k}_p>k_p$ and $k_{p+1}<\widehat{k}_{p+1}$, then

$$\begin{aligned}&\frac{1}{\sqrt{2\pi n}}\sum_{j=\widehat{k}_p+1}^{\widehat{k}_{p+1}}(m_j-\widehat{X}_j^{(R-1)})e^{-ij\omega_l} \\ &= \frac{1}{\sqrt{2\pi n}}\Bigg\{-\Delta_{p+1}\frac{\widehat{k}_{p+1}-k_{p+1}}{\widehat{k}_{p+1}-\widehat{k}_p}\sum_{j=\widehat{k}_p+1}^{k_{p+1}} e^{-ij\omega_l}\end{aligned}$$

$$+ \Delta_{p+1}\frac{k_{p+1}-\widehat{k}_p}{\widehat{k}_{p+1}-\widehat{k}_p}\sum_{j=k_{p+1}+1}^{\widehat{k}_{p+1}} e^{-ij\omega_l} - \frac{1}{\widehat{k}_{p+1}-\widehat{k}_p}\sum_{j=\widehat{k}_p+1}^{\widehat{k}_{p+1}} \epsilon_j \sum_{j=\widehat{k}_p+1}^{\widehat{k}_{p+1}} e^{-ij\omega_l}\Bigg\}$$
$$= O_p\left(\frac{1}{\sqrt{n\Delta_{p+1}^2 l^2}} + \frac{1}{\sqrt{n\Delta_{p+1}^2}} + \frac{1}{l}\right) = O_p\left(\frac{1}{\sqrt{n\Delta_{p+1}^2}} + \frac{1}{l}\right) \tag{A.7}$$

from the assumptions (A1)-(A4). The asymptotic orders for the other cases, for example when $\widehat{k}_p \le k_p$ and $k_{p+1} \ge \widehat{k}_{p+1}$, are exactly the same, and will not be derived here. Hence,

$$x_5 = O_p\left(\sum_{p\in\mathcal{P}\setminus\{i-1\}} \frac{1}{\sqrt{n\Delta_{p+1}^2}} + \frac{1}{l}\right).$$

By factoring out $n\Delta_i^2$,

$$|l^2\nu_l| = O_p\left(\sqrt{\frac{m^2}{n\Delta_i^2}}\right)$$

uniformly in $l = 1, \dots, m$. Therefore, (A.5) follows from

$$\sup_{\Theta_1\le d\le\Theta_2}\left|\sum_{l=1}^{m} l^{2d-2} l^2 \nu_l\right| = o_p(1)$$

because of the assumption (A5) and the absolute summability of $\sum_{j=1}^{\infty} j^{2d-2}$ at $d = \Theta_2$.

Note now that

$$0 < \sum_{l=1}^{\infty} l^{2d-2}(c_1(l)^2 + c_2(l)^2) < \infty.$$

Since $1/m\sum_{l=1}^{m}\log l \to \infty$, it follows from (A.4) that

$$\widehat{d}_{lw} \xrightarrow{p} \Theta_2. \tag{A.8}$$

Finally, from integration by parts,

$$P(\mathcal{N}(0,1/4) > z) < \frac{1}{\sqrt{8\pi}}\frac{1}{z}e^{-2z^2} < e^{-2z^2}$$

for sufficiently large z. This implies that for a given size of test

$$\alpha(n) = P(\mathcal{N}(0,1/4) > c(n)),$$

the critical value $c(n) < \sqrt{-.5\log\alpha(n)}$. In particular, if $\alpha(n) = K/n$ for some positive constant K, then $c(n) < \sqrt{-.5\log(K/n)}$. Therefore, by using (A.8) and the assumption (A5),

$$P(\sqrt{m}\widehat{d}_{lw} > c(n)) \to 1$$

for $\alpha(n) \to 0$ but $\liminf_{n\to\infty} n\alpha(n) > 0$. Therefore, $P(\widehat{R} < R) \to 0$ as the sample size increases.

On the other hand, when $\widehat{R} > R$, we will argue that when testing $\mathrm{H}_0 : \mathrm{CM}\text{-}R$,

$$P(\sqrt{m}\widehat{d}_{lw} > c(n)) \to 0, \tag{A.9}$$

where $c(n)$ is the critical value of $\mathcal{N}(0, 1/4)$ with significance level $\alpha(n)$. Thus, the null hypotheses is rejected at the R-th stage with probability tending to 0, yielding $P(\widehat{R} > R) \to 0$. The result (A.9) immediately follows by showing that at the R-th stage and under $\mathrm{H}_0 : \mathrm{CM}\text{-}R$,

$$\sqrt{m}\widehat{d}_{lw} \xrightarrow{d} \mathcal{N}(0, 1/4), \tag{A.10}$$

since $c(n) \to \infty$ as $\alpha(n) \to 0$.

To show (A.10), let $\widehat{k}_1, \widehat{k}_2, \dots, \widehat{k}_R$ be the R estimated break points at the R-th stage, and set $\widehat{k}_0 = 1$, $\widehat{k}_{R+1} = n$. Then, the periodogram based on $R_j^{(R)} = X_j - \widehat{X}_j^{(R)}$ becomes

$$I_R(\omega_l) = \frac{1}{2\pi n}\left|\sum_{j=1}^{n} \epsilon_j e^{-ij\omega_l} + +\sum_{p=0}^{R}\sum_{j=\widehat{k}_p+1}^{\widehat{k}_{p+1}} (m_j - \widehat{X}_j^{(R)})e^{-ij\omega_l}\right|^2. \tag{A.11}$$

Since the true model is CM-R, the assumption (A4) implies that all true break point fractions are estimated consistently. Therefore, as argued in (A.7), one can verify that (A.11) can be written as

$$I_R(\omega_l) = |x_1|^2 + \nu_l, \quad \nu_l = O_p\left(\sum_{p=1}^{R} \frac{1}{\sqrt{n\Delta_p^2}} + \frac{1}{l}\right)$$

uniformly in $l = 1, 2, \dots, n$. The asymptotic normality in (A.10) follows by using the same argument as in the proof of Theorem 3 of Baek and Pipiras (2012) where a single break point is considered (along the results of Robinson (1995a)). The basic idea is that ν_l is negligible since

$$\frac{m \log^2 m}{n\Delta_p^2} \to 0$$

for all $p = 1, \dots, R$ which follows from the assumption (A5). □

Appendix B: Proof of Theorem 3.3

We assume the following conditions throughout the section.

Assumptions

(B1) Assumptions of Theorem 1 in Robinson (1995a) hold for the time series $\{X_n\}$ with the true LRD parameter $d^* \in (0, 1/2)$ and the number of low frequencies m used in the LW estimation.

(B2) Suppose that

$$\frac{1}{n^{d^*+1/2}} \sum_{1 \le j \le nt} (X_j - EX_j) \xrightarrow{d} \sigma B_{d^*+1/2}(t) \quad \text{in } D[0,1], \tag{B.1}$$

where $\sigma > 0$ and $B_{d^*+1/2}$ is a standard fractional Brownian motion.

Proof of Theorem 3.3. Let $\widehat{d}_{lw}^R$ be the LW estimator of the LRD parameter after removing R breaks where the R estimated break points, $\widehat{k}_1, \dots, \widehat{k}_R$, are obtained through the LSE method in a sequential way. Set also $\widehat{k}_0 = 0$ and $\widehat{k}_{R+1} = n$. Then, (3.14) follows from

$$\widehat{d}_{lw}^R \xrightarrow{p} d^*, \tag{B.2}$$

since $\mathrm{H}_0 : \text{CM-}R$ is rejected for any R as argued following (A.8) as long as $\alpha(n)$ is such that $\alpha(n) \to 0$ but $\liminf_{n\to\infty} n\alpha(n) > 0$.

Observe that, after removing R break points, the periodogram becomes

$$\begin{aligned} I_R(\omega_l) &= \left| \frac{1}{\sqrt{2\pi n}} \sum_{j=1}^{n} (X_j - \widehat{X}_j^{(R)}) e^{-ij\omega_l} \right|^2 \\ &= \left| \frac{1}{\sqrt{2\pi n}} \sum_{j=1}^{n} X_j e^{-ij\omega_l} - \frac{1}{\sqrt{2\pi n}} \sum_{p=0}^{R} \overline{X}_p \sum_{j=\widehat{k}_p+1}^{\widehat{k}_{p+1}} e^{-ij\omega_l} \right|^2 = I_X(\omega_l) + \nu_l, \end{aligned} \tag{B.3}$$

where

$$\overline{X}_p = \frac{1}{\widehat{k}_{p+1} - \widehat{k}_p} \sum_{j=\widehat{k}_p+1}^{\widehat{k}_{p+1}} X_j$$

is the sample average on the p-th regime $(\widehat{k}_p, \widehat{k}_{p+1}]$,

$$I_X(\omega_l) = \left| \frac{1}{\sqrt{2\pi n}} \sum_{j=1}^{n} X_j e^{-ij\omega_l} \right|^2 =: |\kappa(\omega_l)|^2$$

is the periodogram of LRD series $\{X_n\}$ with parameter $d^* \in (0, 1/2)$, and ν_l denotes remaining terms in the expansion.

The convergence (B.2) can be proved by the same argument as Theorem 4 in Baek and Pipiras (2012) where a single break is considered. The key idea is that the periodogram $I_R(\omega_l)$ is dominated by $I_X(\omega_l)$, and ν_l is negligible. Here, we will only show that the asymptotic order of ν_l is the same as that in the proof of Theorem 4 of Baek and Pipiras (2012) even if we subtract multiple local mean levels.

First, from the equation (3.16) in Robinson (1995a), note that

$$\kappa(\omega_l) = O_p(g_l^{1/2}) = O_p\left(\omega_l^{-d^*}\right) \tag{B.4}$$

uniformly in $l = 1, \ldots, m$, where $g_l = G_0 \omega_l^{-2d^*}$ with $G_0 > 0$ appearing in the spectral density of X, $f_X(\omega) \sim G_0 \omega^{-2d^*}$ as $\omega \to 0$. Furthermore, note that

$$\sum_{j=\widehat{k}_p+1}^{\widehat{k}_{p+1}} e^{-ij\omega_l} = O_p\left(\frac{1}{|1-e^{-i\omega_l}|}\right) = O_p\left(\frac{1}{\omega_l}\right). \tag{B.5}$$

To calculate the order of the sample average $\overline{X}_p$ on the p-th regime, rewrite the break points as

$$\widehat{k}_{(1)}, \widehat{k}_{(2)}, \ldots, \widehat{k}_{(R)}$$

according to their sequential order. Observe from (B.1) that

$$\begin{aligned}
&\frac{1}{n^{d^*}} \operatorname*{argmax}_{1\le k\le n-1} \left| \left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^{-1/2} \frac{1}{\sqrt{n}} \left(\sum_{j=1}^{k} X_j - \frac{k}{n}\sum_{j=1}^{n} X_j\right)\right| \\
&\quad \xrightarrow{d} \operatorname*{argsup}_{0<t<1} \left| \big(t(1-t)\big)^{-1/2} \sigma\big(B_{d^*+1/2}(t) - tB_{d^*+1/2}(1)\big)\right| \\
&\quad = \operatorname*{argsup}_{0<t<1} \left| \big(t(1-t)\big)^{-1/2} \sigma \mathcal{W}_{d^*+1/2}(t)\right|,
\end{aligned}$$

where $\mathcal{W}_{d^*+1/2} = B_{d^*+1/2}(t) - tB_{d^*+1/2}(1)$. Therefore, in view of (2.6),

$$\frac{\widehat{k}_{(1)}}{n} \xrightarrow{d} \xi_{(1)} := \operatorname*{argsup}_{0<t<1} \left| \big(t(1-t)\big)^{-1/2} \sigma \mathcal{W}_{d^*+1/2}(t)\right|. \tag{B.6}$$

Similarly by using (B.1) and (B.6), if $\widehat{k}_{(2)} < \widehat{k}_{(1)}$, then

$$\begin{aligned}
&\frac{1}{n^{d^*}} \operatorname*{argmax}_{1\le k<\widehat{k}_{(1)}} \left| \left(\frac{k}{\widehat{k}_{(1)}}\left(1-\frac{k}{\widehat{k}_{(1)}}\right)\right)^{-1/2} \frac{1}{\sqrt{\widehat{k}_{(1)}}} \left(\sum_{j=1}^{k} X_j - \frac{k}{\widehat{k}_{(1)}}\sum_{j=1}^{\widehat{k}_{(1)}} X_j\right)\right| \\
&\quad \xrightarrow{d} \operatorname*{argsup}_{0<t<\xi_{(1)}} \left| \left(\frac{t}{\xi_{(1)}}\left(1-\frac{t}{\xi_{(1)}}\right)\right)^{-1/2} \frac{\sigma}{\sqrt{\xi_{(1)}}} \left(B_{d^*+1/2}(t) - \frac{t}{\xi_{(1)}} B_{d^*+1/2}(\xi_{(1)})\right)\right|
\end{aligned}$$

and if $\widehat{k}_{(2)} > \widehat{k}_{(1)}$,

$$\begin{aligned}
&\frac{1}{n^{d^*}} \operatorname*{argmax}_{\widehat{k}_{(1)}<k<n} \left| \left(\frac{k-\widehat{k}_{(1)}}{n-\widehat{k}_{(1)}}\left(1-\frac{k-\widehat{k}_{(1)}}{n-\widehat{k}_{(1)}}\right)\right)^{-1/2} \right. \\
&\quad \left. \times \frac{1}{\sqrt{n-\widehat{k}_{(1)}}} \left(\sum_{j=\widehat{k}_{(1)}+1}^{k} X_j - \frac{k-\widehat{k}_{(1)}}{n-\widehat{k}_{(1)}} \sum_{j=\widehat{k}_{(1)}+1}^{n} X_j\right)\right|
\end{aligned}$$

$$\xrightarrow{d} \operatorname*{argsup}_{\xi_{(1)}<t<1} \Bigg| \left(\frac{t-\xi_{(1)}}{1-\xi_{(1)}} \left(1 - \frac{t-\xi_{(1)}}{1-\xi_{(1)}} \right) \right)^{-1/2} \frac{\sigma}{\sqrt{1-\xi_{(1)}}}$$
$$\times \left(B_{d^*+1/2}(t) - B_{d^*+1/2}(\xi_{(1)}) - \frac{t-\xi_{(1)}}{1-\xi_{(1)}} (B_{d^*+1/2}(1) - B_{d^*+1/2}(\xi_{(1)})) \right) \Bigg|.$$

This shows that $\widehat{k}_{(2)}/n$ converges in distribution. Proceeding similarly, one can show that

$$\frac{\widehat{k}_{(i)}}{n} \xrightarrow{d} \xi_{(i)}, \tag{B.7}$$

for $i = 1, 2, \ldots, R$. A similar argument also yields that

$$n^{1/2-d^*} \overline{X}_p = \frac{n^{1/2-d^*}}{\widehat{k}_{p+1} - \widehat{k}_p} \sum_{j=\widehat{k}_p+1}^{\widehat{k}_{p+1}} X_j \xrightarrow{d} \frac{1}{\xi_{p+1} - \xi_p} \left(B_{d^*+1/2}(\xi_{p+1}) - B_{d^*+1/2}(\xi_p) \right),$$

and hence that

$$\overline{X}_p = O_p(n^{d^*-1/2}). \tag{B.8}$$

Therefore, from (B.4), (B.5) and (B.8), the order of ν_l becomes

$$\nu_l = O_p \left(\frac{n^{2d^*}}{l^{d^*+1}} \right), \tag{B.9}$$

which is exactly the same order as in the proof of Theorem 4 in Baek and Pipiras (2012). □

Acknowledgements

We would like to thank the Editor and anonymous Referees for many useful comments and suggestions.

References

Abadir, K., Distaso, W. and Giraitis, L. (2007). Nonstationarity-extended local Whittle estimation. *Journal of Econometrics* **141** 1353–1384. MR2413505

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59** 817–858. MR1106513

Andrews, D. W. K. and Sun, Y. (2004). Adaptive local polynomial Whittle estimation of long-range dependence. *Econometrica* **72** 569–614. MR2036733

Baek, C. and Pipiras, V. (2012). Statistical tests for a single change in mean against long-range dependence. *Journal of Time Series Analysis* **33** 131–151. MR2877613

Bai, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory* **13** 315–352. MR1455175

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78. MR1616121

Beran, J. (1994). *Statistics for Long-Memory Processes. Monographs on Statistics and Applied Probability* **61**. Chapman and Hall, New York. MR1304490

Berkes, I., Horváth, L., Kokoszka, P. and Shao, Q.-M. (2006). On discriminating between long-range dependence and changes in mean. *The Annals of Statistics* **34** 1140–1165. MR2278354

Bisaglia, L. and Gerolimetto, M. (2009). An Empirical Strategy to Detect Spurious Effects in Long Memory and Occasional-Break Processes. *Communications in Statistics - Simulation and Computation* **38** 172–189. MR2518339

Box-Steffensmeier, J. M. and Tomlinson, A. R. (2000). Fractional integration methods in political science. *Electoral Studies* **19** 63–76.

Byers, D., Davidson, J. and Peel, D. (1997). Modelling political popularity: An analysis of long-range dependence in opinion poll series. *Journal of the Royal Statistical Society. Series A: Statistics in Society* **160** 471–490.

Chang, S. and Perron, P. (2014). A comparison of alternative methods to construct confidence intervals for the estimate of a break date in linear regression models Preprint.

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. *Wiley Series in Probability and Statistics*. John Wiley & Sons Ltd., Chichester. MR2743035

Diebold, F. X. and Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics* **105** 131–159. MR1864215

Dmowska, R. and Saltzman, B. (1999). *Long-Range Persistence in Geophysical Time Series*. Academic Press, London, UK.

Doukhan, P., Oppenheim, G. and Taqqu, M. S. (2003). *Theory and Applications of Long-Range Dependence*. Birkhäuser Boston Inc., Boston, MA. MR1956041

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and its Application*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York. Probability and Mathematical Statistics. MR0624435

Henry, M. (2001). Robust automatic bandwidth for long memory. *Journal of Time Series Analysis* **22** 293–316. MR1837370

Hušková, M. and Kirch, C. (2008). Bootstrapping confidence intervals for the change-point of time series. *Journal of Time Series Analysis* **29** 947–972. MR2464948

Iacone, F. (2010). Local Whittle estimation of the memory parameter in presence of deterministic components. *Journal of Time Series Analysis* **31** 37–49. MR2640845

Jach, A. and Kokoszka, P. (2008). Wavelet-domain test for long-range dependence in the presence of a trend. *Statistics. A Journal of Theoretical and Applied Statistics* **42** 101–113. MR2414352

Juhl, T. and Xiao, Z. (2009). Tests for changing mean with monotonic power. *Journal of Econometrics* **148** 14–24. MR2494814

Klemeš, V. (1974). The Hurst pheomenon: a puzzle? *Water Resources Research* **10** 675–688.

Kuan, C.-M. and Hsu, C.-C. (1998). Change-point estimation of fractionally integrated processes. *Journal of Time Series Analysis* **19** 693–708. MR1665953

Kunsch, H. R. (1989). The Jackknife and the Bootstrap for General Stationary Observations. *The Annals of Statistics* **17** 1217–1241. MR1015147

Lahiri, S. N. (1999). Theoretical Comparisons of Block Bootstrap Methods. *The Annals of Statistics* **27** 386–404. MR1701117

Lavielle, M. (2005). Using penalized contrasts for the change-point problem. *Signal Processing* **85** 1501–1510.

Lebo, M. J., Walker, R. W. and Clarke, H. (2000). You must remember this: Dealing with long memory in political analyses. *Electoral Studies* **19** 31–48.

Liu, J., Wu, S. and Zidek, J. V. (1997). On segmented multivariate regression. *Statistica Sinica* **7** 497–525. MR1466692

McCloskey, A. and Perron, P. (2013). Memory parameter estimation in the presence of level shifts and deterministic trends. *Econometric Theory* **29** 1196–1237. MR3148830

Mikosch, T. and Stărică, C. (2004). Changes of structure in financial time series and the GARCH model. *REVSTAT Statistical Journal* **2** 41–73. MR2259345

Mills, T. C. (2007). Time series modelling of two millennia of northern hemisphere temperatures: long memory or shifting trends? *Journal of the Royal Statistical Society. Series A. Statistics in Society* **170** 83–94. MR2339099

Ohanissian, A., Russell, J. R. and Tsay, R. S. (2008). True or Spurious Long Memory? A New Test. *Journal of Business & Economic Statistics* **26** 161–175. MR2420145

Palma, W. (2007). *Long-Memory Time Series*. John Wiley & Sons, Inc., Hoboken, New Jersey, USA. MR2297359

Park, K. and Willinger, W. (2000). *Self-Similar Network Traffic and Performance Evaluation*. John Wiley & Sons, Inc., New York, NY, USA.

Perron, P. (2006). Dealing with Structural Breaks. In *Palgrave Handbook of Econometrics* 278–352. Palgrave Macmillan, London.

Qu, Z. (2011). A test against spurious long memory. *Journal of Business and Economic Statistics* **29** 423–438. MR2848513

Robbins, M., Gallagher, C., Lund, R. and Aue, A. (2011). Mean shift testing in correlated data. *Journal of Time Series Analysis* **32** 498–511. MR2835683

Robinson, P. M. (1995a). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics* **23** 1630–1661. MR1370301

Robinson, P. M. (1995b). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics* **23** 1048–1072. MR1345214

Robinson, P. M. (2003). *Time Series with Long Memory*. Oxford University Press, Oxford. MR2083220

Robinson, P. M. (2005). Robust covariance matrix estimation: HAC estimates with long memory/antipersistence correction. *Econometric Theory* **21** 171–180. MR2153861

Roughan, M. and Veitch, D. (1999). Measuring long-range dependence under changing traffic conditions. In *Proceedings of IEEE INFOCOM* **3** 1513–1521.

Seijo, E. and Sen, B. (2011). Change-point in stochastic design regression and the bootstrap. *The Annals of Statistics* **39** 1580–1607. MR2850213

Sen, A. and Srivastava, M. S. (1975). On tests for detecting change in mean when variance is unknown. *Annals of the Institute of Statistical Mathematics* **27** 479–486. MR0415864

Stărică, C. and Granger, C. (2005). Nonstationarities in Stock Returns. *The Review of Economics and Statistics* **87** pp. 503–522.

Yao, Y.-C. (1988). Estimating the number of change-points via Schwarz' criterion. *Statistics & Probability Letters* **6** 181–189. MR0919373

Yau, C. Y. and Davis, R. A. (2012). Likelihood inference for discriminating between long-memory and change-point models. *Journal of Time Series Analysis* **33** 649–664. MR2944844