Change-in-Mean Tests in Long-memory Time Series: A Review of Recent Developments*

Kai Wenger[a] Christian Leschinski[a] Philipp Sibbertsen[a]

[a]Leibniz University Hannover

June 15, 2017

Abstract

It is well known that standard tests for a mean shift are invalid in long-range dependent time series. Therefore, several long memory robust extensions of standard testing principles for a change-in-mean have been proposed in the literature. These can be divided into two groups: those that utilize consistent estimates of the long-run variance and self-normalized test statistics.

Here, we review this literature and complement it by deriving a new long memory robust version of the sup-Wald test. Apart from giving a systematic review, we conduct an extensive Monte Carlo study to compare the relative performance of these methods. Special attention is paid to the interaction of the test results with the estimation of the long-memory parameter. Furthermore, we show that the power of self-normalized test statistics can be improved considerably by using an estimator that is robust to mean shifts.

Key words: Fractional Integration · Structural Breaks · Long Memory.

JEL classification: C12, C22

1 Introduction

Both long memory and mean shifts generate similar time series features such as significant autocorrelations at large lags or a pole in the periodogram at Fourier frequencies local to zero (see for example Diebold and Inoue (2001), Granger and Hyung (2004), Mikosch and Stărică (2004)). If long memory is falsely detected in a short-memory time series contaminated by a mean shift, it is referred to as 'spurious long memory'. Tests to distinguish spurious from true long memory are, for example, proposed by Ohanissian, Russell, and Tsay (2008) and Qu (2011).

However, this ambiguity is not unidirectional. Traditional testing techniques for a mean shift are similarly invalidated by long memory in the time series. The problem is that the standard tests are developed for independent or serially correlated innovations. They do not allow for

*Financial support of the Deutsche Forschungsgesellschaft (DFG) is gratefully acknowledged.

long-range dependence. Therefore, the limiting distributions of the tests are different and the normalization factors are too small. It can be shown that these tests reject the null hypothesis of a constant mean with probability of one, asymptotically, if long memory is present in the series (see Wright (1998) and Krämer and Sibbertsen (2002), among others). The problem is circular. Even if it is suspected that a structural break has taken place, and one therefore wants to test this hypothesis, a series with a mean shift will always show features of long-memory processes. The presence of long memory can therefore never be ruled out with certainty.

To circumvent this issue, it is necessary to use long memory robust tests against mean shifts. This is why numerous studies modify existing standard change-in-mean tests such as CUSUM tests originally proposed by Brown, Durbin, and Evans (1975), Wilcoxon-type rank tests (e.g. Bauer (1972)), and sup-Wald type tests by Andrews (1993) to account for long memory.

An early attempt was made, for example, by Hidalgo and Robinson (1996) who modify the sup-Wald test to be valid under long-memory, but under the assumption that the change point is known. For breaks at an unknown point in time Horváth and Kokoszka (1997) propose a modified CUSUM test statistic for Gaussian processes that uses a different normalization factor and long-run variance under long memory. They show that it converges to a fractional Brownian bridge. Therefore, both the normalization and the critical values depend on the memory parameter d. Wang (2008) extends the results of Horváth and Kokoszka (1997) to a wider range of linear processes and shows that one can use estimators of the long-run variance and d asymptotically. Dehling, Rooch, and Taqqu (2013) derive the limiting distribution of a Wilcoxon-type test with an appropriate long run variance estimator for $d > 0$.

All these methods have in common that they follow the traditional approach to employ a consistent estimate of the long-run variance for the standardization of the test statistic. Another series of contributions replaces this estimate by an inconsistent self-normalization, as advocated for weakly correlated time series by Shao and Zhang (2010).

In the context of testing on a mean shift in long-memory time series the technique is applied in the sup-Wald test of Shao (2011). Betken (2016) also uses self-normalization to standardize a Wilcoxon-type test. Later, Iacone, Leybourne, and Taylor (2014) state that the self-normalization is a special case of fixed-b estimation of the long-run variance, as proposed by Kiefer and Vogelsang (2005) for short-memory processes and generalized by McElroy and Politis (2012) to long-memory processes. Iacone, Leybourne, and Taylor (2014) then introduce a sup-Wald test on a change-in-mean under long memory based on a fixed-b estimator of the long-run variance.

Here, we give a detailed review of this literature and close a gap by deriving a new long memory robust version of the sup-Wald test of Andrews (1993). Our test falls into the first group that utilizes a consistent estimate of the long-run variance. In this case it is based on regression with long-memory errors as considered in Yajima (1988).

The main contribution, however, is made by means of an extensive Monte Carlo study that we use to evaluate the relative performance of all the aforementioned methods in finite samples. So far most of the contributions have made simulations for the proposed test alone, or pairwise comparisons with only one of the previously available methods. It is therefore not possible to make any recommendations for the selection of methods in practice. In regard of the extent of the literature, and the recent surge of interest in this topic, the comparison made here will be helpful as a reference for further research and as guidance for practical applications.

Previous studies have also mostly abstracted from the effect of estimating the memory parameter that appears in the test statistics and/or the limit distributions. This is a crucial simplification since d is not known in practice. Hence, we pay special attention to the interaction of the test results with the estimation of d. Furthermore, we show that the power of self-normalized test statistics can be improved considerably by using an estimator that is consistent under the null hypothesis as well as under the alternative of a mean shift. A modified local Whittle estimator with these properties was proposed by Hou and Perron (2014).

The rest of the paper is organized as follows. Section 2 gives the hypothesis of interest and describes the first group of tests on change-in-mean under long memory with consistent normalizations. The second group that utilizes self-normalization is considered in Section 3. In Section 4 we discuss different estimation methods for the long-memory parameter. Section 5 presents the Monte-Carlo simulation and makes recommendations for empirical applications. Finally, Section 6 concludes.

2 Change-in-mean tests under long memory

2.1 Model

We consider a single mean shift model where the stochastic process $(y_t)_{t\geq 1}$ is generated by

$$y_t = \mu + \beta \mathbb{1}(t \geq t^*) + \epsilon_t. \tag{1}$$

Here, the sequence of regression means $\mu_t = \mu + \beta \mathbb{1}(t \geq t^*)$ for $t \geq 0$ is assumed to be deterministic and fulfills $|\mu_t| < \infty$. The coefficient β determines the magnitude of the possible mean shift, $\mathbb{1}(t \geq t^*)$ is an indicator function that depends on the location of the break point denoted by $t^* = \lfloor \tau^* T \rfloor$, where $\lfloor \cdot \rfloor$ returns the integer part of its argument and $\tau^* \in (0,1)$ denotes the break fraction. Finally, $(\epsilon_t)_{t\geq 1}$ is a stationary fractionally integrated error term $\epsilon_t = (1-L)^{-d} v_t$ with long memory parameter $|d| < 1/2$ and v_t is a stationary short memory sequence with continuous and bounded spectrum and infinite past. The error term $(\epsilon_t)_{t\geq 1}$ is therefore a fractionally integrated process of type I (cf. Marinucci and Robinson (1999)).

As usual,

$$(1-L)^d = \sum_{k=0}^{\infty} \pi_k(d) L^k = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)} L^k \tag{2}$$

is the fractional differencing operator and $\Gamma(\cdot)$ denotes the Gamma function. For $d \in (-0.5, 0) \bigcup (0, 0.5)$ the sequence ϵ_t satisfies

$$\begin{aligned} \gamma(k) &\sim L_\gamma(k) |k|^{2d-1} \quad \text{for} \quad k \to \infty \;, \\ L_\gamma(k) &= 2G(k^{-1}) \Gamma(1-2d) \sin(\pi d) \;, \end{aligned} \tag{3}$$

where $\gamma(k)$ is the autocovariance function evaluated at lag k, and $G(\cdot)$ is a slowly varying function in Zygmund's sense (such that for $k \to \infty$ the function $G(\cdot)$ is well approximated by a constant G).

Now, given a specific time series $y_1, ..., y_T$ the interest lies on testing the null hypothesis of a constant unconditional mean

$$H_0 : \mu_1 = ... = \mu_T = \mu$$

against the alternative of a change-in-mean

$$H_1 : \mu_t \neq \mu_s \quad \text{for some} \quad 1 < t, s < T,$$

which is equivalent to test the null hypothesis $H_0 : \beta = 0$ against the alternative $H_1 : \beta \neq 0$ in (1).

2.2 CUSUM tests under long memory

The standard CUSUM test rejects the null hypothesis of a constant mean if the correctly standardized partial sum of the estimated residuals from a linear regression is too high. Krämer and Sibbertsen (2002), among others, show that the standardization of the CUSUM test is too small and a different long-run variance estimator has to be used when highly persistent long-memory time series are considered. Furthermore, the wrong limiting distribution is assumed. Together this leads to an asymptotic rejection probability of one under H_0 such that the test is not applicable.

Hence, Horváth and Kokoszka (1997) modified the CUSUM test statistic for Gaussian processes, which is extended by Wang (2008) to general linear processes. Both of these contributions replace the wrong scaling parameter as well as the long-run variance estimator in the standard method for $d > 0$ and derive the limiting distribution of the modified test. For $|d| < 0.5$ the test is based on

$$S_T^{CLM}(t^*, d) = \frac{1}{\widehat{\sigma} T^{1/2+d}} \sum_{t=1}^{t^*} \widehat{\epsilon}_t, \tag{4}$$

where $\widehat{\epsilon}_t$ is an estimate of the residuals from (1), and $\widehat{\sigma}^2$ is a consistent estimator of the long-run variance under long memory. This can be, for example, the memory and autocorrelation consistent (MAC) estimator of Robinson (2005) and Abadir, Distaso, and Giraitis (2009). The long-run variance is given by

$$\sigma^2 = \begin{cases} \frac{\Gamma(1-2d) G \, 2\sin(\pi d)}{d(1+2d)}, & \text{for } d \in (-0.5, 0) \bigcup (0, 0.5), \\ 2\pi G, & \text{for } d = 0, \end{cases}$$

where $G \in (0, \infty)$ the constant from (3). The CUSUM under long memory (CUSUM-LM) test statistic is then given by

$$Q_T^{CLM} = \sup_{t^* \in (t_1^*, ..., t_2^*)} |S_T^{CLM}(t^*, d)|$$

and converges in distribution to a fractional Brownian bridge

$$Q_T^{CLM} \Rightarrow \sup_{\tau \in [\tau_1, \tau_2]} |B_d(\tau) - \tau B_d(1)|$$

where $B_d(\tau)$ is a fractional Brownian motion of type I. Convergence in distribution is denoted by " $\Rightarrow$ ". It is usual to restrict the search area to some interval $(t_1^*, ..., t_2^*) = (\lfloor T\tau_1 \rfloor, ..., \lfloor T\tau_2 \rfloor)$ in which the statistic is calculated. Andrews (1993) suggests using $\tau_1 = 1 - \tau_2 = 0.15$.
The standard CUSUM test for short memory time series ($d = 0$) is nested in (4). In this case the MAC estimator $\widehat{\sigma}^2$ is replaced with a heteroscedasticity and autocorrelation consistent (HAC) long-run variance estimator (cf. for example Newey and West (1987), and Andrews (1991)). The limit distribution then simplifies to the supremum of a standard Brownian bridge.

2.3 Wilcoxon-type tests under long memory

Given a time series $y_1, ..., y_T$ the non-parametric Wilcoxon two-sample rank test divides the sample at some potential break point t^* in two parts. For each observation in the first part, it sums over the number of times for which it is smaller than the observations in the second part.
An adaption of the standard test to long-memory time series is made by Dehling, Rooch, and Taqqu (2013) based on

$$S_T^{WLM}(t^*, d) = \frac{1}{T\Theta_T} \left| \sum_{i=1}^{t^*} \sum_{j=t^*+1}^{T} \left(\mathbb{1}(y_i \leq y_j) - \frac{1}{2} \right) \right|. \tag{5}$$

Here, Θ_T is a scaling function given by

$$\Theta_T^2 = \frac{L_\gamma T^{1+2d}}{d(1+2d)}, \tag{6}$$

and L_γ is the slowly varying function from (3). The Wilcoxon long memory (WLM) test then rejects for large values of the test statistic

$$Q_T^{WLM} = \max_{t^* \in (t_1^*, ..., t_2^*)} |S_T^{WLM}(t^*, d)|.$$

For practical implementation L_γ has to be estimated. In case of fractional white noise we have $\Theta_T = T^{1/2+d}$. For $d \in (0, 1/2)$ Dehling, Rooch, and Taqqu (2013) show that the test statistic converges in distribution to a scaled fractional Brownian Bridge

$$Q_T^{WLM} \Rightarrow \frac{1}{2\sqrt{\pi}} \sup_{\tau \in [\tau_1, \tau_2]} |B_d(\tau) - \tau B_d(1)|,$$

which is identical to the limiting distribution of the CUSUM-LM test - except for the multiplicative constant. One disadvantage of the test is that it does not allow for $d \leq 0$. Under the assumption that d is known Dehling, Rooch, and Taqqu (2013) compare the WLM test with the CUSUM-LM test and find that the WLM delivers better results for heavy tailed data, but not for fractionally integrated white noise.

2.4 Sup-Wald tests under long memory

The idea of the sup-Wald test is to perform a t-test on the estimate of β in the single mean shift model (1) for each possible break date t^*. Wright (1998) shows that the standard sup-Wald test diverges under long memory. Here, as a new testing procedure, we suggest using a modified sup-Wald test that allows for long-memory errors as in Yajima (1988) and uses the corresponding estimate for the long-run variance in the denominator.

This approach closes a gap in the literature. While the long memory robust extensions of the CUSUM and Wilcoxon tests (discussed above) employ consistent estimates of the long-run variance, the existing sup-Wald tests under long memory (discussed in the next section) all use self-normalizations.

With $x_t(t^*) := (1, \mathbb{1}_t(t^*))'$ and using the OLS estimate $\widehat{\beta}(t^*)$ the test is based on

$$S_T^{SWLM}(t^*, d) = \frac{\widehat{\beta}(t^*)^2}{\widehat{\sigma}^2 c M_{xx}(t^*)^{-1} c' T^{2d}}, \tag{7}$$

where $c = (0, 1)$, $M_{xx}(t^*) = \sum_{t=1}^{T} x_t(t^*) x_t(t^*)'$, and $\widehat{\sigma}^2$ is the MAC long-run variance estimator of Robinson (2005). Our sup-Wald test statistic under long memory (sup-Wald-LM) is then given by

$$Q_t^{SWLM} = \sup_{t^* \in (t_1^*, \ldots, t_2^*)} |S_T^{SWLM}(t^*, d)|. \tag{8}$$

For this test we obtain the limit distribution stated in the proposition below.

Proposition 1. *For y_t as defined in (1) and Q_t^{SWLM} as in (8), we have*

$$Q_t^{SWLM} \Rightarrow \mathop{sup}_{\tau \in [\tau_1, \tau_2]} V_d(\tau),$$
$$\textit{where} \qquad V_d(\tau) = \frac{(B_d(\tau) - \tau B_d(1))^2}{\tau(1-\tau)}.$$

Proof. *The derivation of the limiting distribution proceeds in direct analogy to that of Theorem 3 in Andrews (1993), but using the long-run variance in regressions with long-range dependent errors derived in Yajima (1988). It is thus omitted.* □

The standard sup-Wald test by Andrews (1993) is nested in the sup-Wald-LM test for $d = 0$. In this case a HAC-estimator can be used in the denominator of the test statistic and the test converges to a functional of standard Brownian motions instead of fractional ones.

3 Self-normalized tests

Instead of using a consistent estimate of the long-run variance in the denominator of the test statistic, the following approaches apply a self-normalization technique which was first introduced by Shao (2010) in the context of inference in time series. Later, Shao and Zhang (2010) considered it for structural change tests, but only in short-memory time series. The self-normalizer is not a consistent estimate of the long-run variance, but proportional to it. By using it, one achieves a

well-defined limiting distribution even when long memory is present in the time series.
The self-normalizer $G(t^*)$ is given by

$$G(t^*) = \left(\frac{1}{T} \sum_{t=1}^{t^*} S_t^2(1, t^*) + \frac{1}{T} \sum_{t=t^*+1}^{T} S_t^2(t^* + 1, T) \right)^{1/2} \tag{9}$$

$$\text{with} \quad S_t(j, t^*) = \sum_{h=j}^{t} (H_h - \overline{H}_{j,t^*}), \quad \text{and} \quad \overline{H}_{j,t^*} = \frac{1}{t^* - j + 1} \sum_{t=j}^{t^*} H_t, \tag{10}$$

where H is the quantity the specific test is based on, for example $H_i = R_i$, for the Wilcoxon test discussed below. In contrast to the consistent long-run variance estimates, one can see that the self-normalization depends on the location of the mean shift and is therefore different for each t^*.

For $T \to \infty$ the self-normalizer $G(t^*)$ converges in distribution to

$$U(\tau, d) = \left(\int_0^\tau J_d^2(s, 0, \tau) ds + \int_\tau^1 J_d^2(s, \tau, 1) ds \right)^{1/2},$$

where $J_d(s, s_1, s_2) = B_d(s) - B_d(s_1) - \frac{s - s_1}{s_2 - s_1}(B_d(s_2) - B_d(s_1))$, for $s \in [s_1, s_2]$, and $0 < s_1 < s_2 < 1$. Note that the self-normalizer $G(t^*)$ does not explicitly account for the memory d. Instead, the dependence of the quantity on d is subsumed under the limit distribution $U(\tau, d)$, alone.

3.1 Self-normalized Wilcoxon tests

First, we consider the self-normalized Wilcoxon-type change point test by Betken (2016). It modifies the WLM test from (5) by replacing the denominator with the self-normalizer in (9), where the quantity of interest H_i is the rank statistic $R_i = \sum_{j=1}^{T} \mathbb{1}(y_j \leq y_i)$. This leads to

$$S_T^{SNW}(t^*) = \frac{\sum_{i=1}^{t^*} R_i - \frac{t^*}{T} \sum_{i=1}^{T} R_i}{G(t^*)}. \tag{11}$$

The numerator of the test statistic performs a CUSUM test in terms of ranks. Asymptotically this is equivalent to the numerator of the WLM test statistic in (5). The self-normalized Wilcoxon (SNW) test is then given by

$$Q_T^{SNW} = \sup_{t^* \in (t_1^*, \ldots, t_2^*)} |S_T^{SNW}(t^*)|.$$

Under H_0 the test statistic converges in distribution to

$$Q_T^{SNW} \Rightarrow \sup_{\tau \in [\tau_1, \tau_2]} \frac{|B_d(\tau) - \tau B_d(1)|}{U(\tau, d)}.$$

Note that the fractional Brownian motions in $U(\tau, d)$ are the same as those appearing in the enumerator. An advantage of the self-normalized version of the test compared to that in (5) is that L_γ in (6) does not need to be estimated.

3.2 Self-normalized sup-Wald tests

Shao (2011) proposes a modified sup-Wald type test where the denominator of the sup-Wald test in (7) is replaced by the self-normalization as in (9), with $H_t = y_t$. The test is based on

$$S_T^{SNSW}(t^*) = \frac{|\widehat{\beta}(t^*)|}{G(t^*)}. \tag{12}$$

Again, the test statistic of the self-normalized sup-Wald (SNSW) test is the supremum of (12) in a user-chosen search area

$$Q_T^{SNSW} = \sup_{t^* \in (t_1^*, \dots, t_2^*)} S_T^{SNSW}(t^*),$$

with limiting distribution

$$Q_T^{SNSW} \Rightarrow \sup_{\tau \in [\tau_1, \tau_2]} \frac{|B_d(\tau)/\tau - \{B_d(1) - B_d(\tau)\}/(1-\tau)|}{U(\tau, d)}.$$

The limiting distribution only depends on the value of the long-memory parameter, but not on nuisance parameters such as the long-run variance. Similar to the limit distribution of the self-normalized Wilcoxon test, the distribution of the self-normalizer appears in the denominator, whereas the enumerator differs due to the different testing principle.

3.3 Fixed-b sup-Wald tests

Nonparametric kernel based HAC estimators consistently estimate the long-run variance in short memory and weakly correlated time series for a user chosen bandwidth $M = \lfloor bT \rfloor$. However, they essentially assume that $b \to 0$ as $T \to \infty$. A different assumption is made by the fixed-b approach (see Kiefer and Vogelsang (2005), Sun, Phillips, and Jin (2008), and Sun (2014), among others), where $b \in (0, 1]$ for $T \to \infty$, i.e. b is held fix. Similar to the self-normalization, the fixed-b estimate is not a consistent estimate for the long-run variance, but converges to the long-run variance multiplied by a functional of a Brownian bridge. It is dependent on the bandwidth and kernel choice and so are the limiting distributions of test statistics using the fixed-b approach. The advantage of using fixed-b is that it delivers more accurate results in finite samples than the traditional HAC estimation.

Using the estimated sample autocovariances $\widehat{\gamma}(t^*) = \frac{1}{T}\sum_{t=j+1}^{T} \widehat{\epsilon}_t(t^*)\widehat{\epsilon}_{t-j}(t^*)$, one obtains the fixed-$b$ long-run variance via

$$\widehat{\sigma}_b^2(t^*) = \widehat{\gamma}_0(t^*) + 2\sum_{j=1}^{T-1} k(j/M)\widehat{\gamma}_j(t^*), \tag{13}$$

which must be estimated for each possible break point t^*. The kernel function $k(j/M)$ and the bandwidth parameter $M = \lfloor bT \rfloor$ are user chosen.

Iacone, Leybourne, and Taylor (2014) modify the standard sup-Wald test by inserting the fixed-b long-run variance estimator $\widehat{\sigma}_b^2(t^*)$ in the denominator of (7) and dropping the term involving the sample size, such that

$$S_T^{FBSW}(t^*) = \frac{\widehat{\beta}(t^*)^2}{\widehat{\sigma}_b^2 c M_{xx}(t^*)^{-1} c'}.$$

The fixed-b sup-Wald (FBSW) test is then given by

$$Q_T^{FBSW} = \sup_{t^* \in (t_1^*, \dots, t_2^*)} S_T^{FBSW}(t^*).$$

They show that Quadratic Spectral, Daniell, Parzen, and Bartlett kernels can be applied, but focus on the Bartlett kernel, that - according to Kiefer and Vogelsang (2005) - delivers the highest power for large b. Under the null hypothesis, Iacone, Leybourne, and Taylor (2014) derive the limiting distribution of the test statistic with $t^*/T \to \tau^*$, as $T \to \infty$. This is given by

$$Q_T^{FBSW} \Rightarrow \sup_{\tau \in [\tau_1, \tau_2]} \frac{V_d(\tau)}{\Phi_d(b, \tau)}.$$

In case of the Bartlett kernel we have

$$\begin{aligned}
\Phi_d(b,\tau) &= \frac{2}{b}\int_0^1 Q_d(s,\tau)^2 ds - \frac{2}{b}\int_0^{1-b} Q_d(s+b,\tau) Q_d(s,\tau) ds, \\
Q_d(s,\tau) &= B_d(s) - (B_d(\tau)/\tau)s - \frac{V_d(\tau)(s-\tau)\mathbb{1}(s>\tau)}{(B_d(\tau) - \tau B_d(1))},
\end{aligned}$$

and $V_d(\tau)$ is defined as in Proposition 1.

Hence, the null limiting distribution is dependent on d, the kernel function, the bandwidth, and the search area. As for the other self-normalized test statistics, the limit distribution is that of the standard test statistic divided by that of the self-normalizier.

Since the SNSW test is based on the sup-Wald principle using the self-normalizer, it can be subsumed under the FBSW tests using a Bartlett kernel with $b = 1$ (cf. Shao (2010)). Iacone, Leybourne, and Taylor (2014) provide simulation results where they find that the FBSW test using the Bartlett kernel and $b = 0.1$ provides better power results than the the SNSW test. This can be explained by the bandwidth choice in fixed-b estimation. Kiefer and Vogelsang (2005) show that the higher the bandwidth, the smaller the size distortions are, but the power decreases as well.

4 Estimation of the memory parameter

The majority of the tests discussed above are derived under the assumption that d is known. Unfortunately, this is not the case in practical applications, where it needs to be estimated. The effect of estimation error in d on the performance of the tests is therefore of major importance. This is one of the main topics in our Monte Carlo simulation.

4.1 Local Whittle estimator

Due to its good asymptotic properties and mild assumptions, the local Whittle estimator by Künsch (1987) and Robinson (1995a) is typically applied to estimate d. For $d \in (-\frac{1}{2}, \frac{1}{2})$, it is given by

$$\widehat{d}_{LW} = \arg\min_d \left[\log \widehat{G}(d) - \frac{2d}{m} \sum_{j=1}^{m} \log \lambda_j \right], \tag{14}$$

where $\widehat{G}(d) = \frac{1}{m} \sum_{j=1}^{m} \lambda_j^{2d} I_{T,y}(\lambda_j)$, $I_{T,y}(\lambda_j) = (2\pi T)^{-1} | \sum_{t=1}^{T} y_t e^{\mathbf{i}t\lambda_j}|^2$ is the periodogram of the series y_t, and the $\lambda_j = (2\pi j/T)$ are the Fourier frequencies. The user-chosen bandwidth m needs to fulfill $1/m + m/T \to \infty$, as $T \to \infty$, to ensure consistency. Under strengthened conditions especially on the bandwidth choice Robinson (1995a) shows that $\sqrt{m}(\widehat{d} - d^0) \sim N(0, 1/4)$, asymptotically. The asymptotic variance of $1/(4m)$ is smaller than that of the log-periodogram estimator of Geweke and Porter-Hudak (1983) and Robinson (1995b), which is $\pi^2/(24m)$. Therefore, the local Whittle estimator is usually preferred.

4.2 Modified local Whittle estimator

As shown by Diebold and Inoue (2001), Granger and Hyung (2004), Mikosch and Stărică (2004), and Perron and Qu (2010), among others, series contaminated by mean shifts and smooth trends display similar characteristics to long memory processes and hence long memory estimation methods such as the local Whittle estimator in (14) are positively biased.

With increasing memory, it becomes easier to confuse mean shifts and long-range dependence. This is why the critical values of the employed tests are monotonic in d. If d is overestimated, the tests therefore lose power, which is always the case under the alternative with standard methods - such as the local Whittle estimator.

We therefore analyze in our Monte Carlo simulation, how the tests perform if d is estimated using the robust estimator of Hou and Perron (2014). They propose a modified local Whittle estimator that is consistent in the presence as well as in the absence of mean shifts, i.e. under the null hypothesis and under the alternative. It is given by

$$\widehat{d}_{mLW} = \arg\min_{d,\theta} \left[\log \widehat{G}(d, \theta) + \frac{1}{m} \sum_{j=1}^{m} \log(\lambda_j^{-2d} + \theta \lambda_j^{-2}/T) \right], \tag{15}$$

where $\widehat{G}(d, \theta) = \frac{1}{m} \sum_{j=1}^{m} \frac{I_{T,y}(\lambda_j)}{\lambda_j^{-2d} + \theta \lambda_j^{-2}/T}$ and θ reflects the scaling of the shifts. Consistency and asymptotic normality are shown by Hou and Perron (2014) for $m > T^{5/9}$.

5 Simulation results

To our knowledge, there is no systematic evaluation of the relative performance of the reviewed change-in-mean tests under long memory. Our main contribution is therefore to provide a comprehensive Monte Carlo simulation study of their finite sample size and power properties that allows to determine which test performs best in which situations.

Our second goal is to investigate the effect of long-memory parameter estimation, which is crucial in practical applications. Therefore, we choose three different scenarios for the long-memory parameter: (i) d is known, (ii) d is estimated by the local Whittle estimator $\widehat{d}_{LW}$, and (iii) d is estimated by the modified local Whittle estimator $\widehat{d}_{mLW}$. We put special emphasis on (ii) and (iii).

As the data generating process we use the following autoregressive fractional integrated processes of type I

$$(1-\phi L)\epsilon_t = (1-L)^{-d} u_t$$

with $\phi \in \{0; 0.5\}$. We choose $d \in \{0; 0.2; 0.4\}$, sample sizes $T \in \{100; 250; 500; 1000\}$, a nominal significance level of $\alpha = 5\%$, and report the rejection frequencies obtained with $M = 5,000$ replications.

Table 1 displays the size results of the introduced tests. On the left hand side of the table, for a known d, and no short run dynamics ($\phi = 0$), almost all tests reach their nominal size. Not surprisingly, the WLM test suffers from substantial size distortions for negative d, since the test is only constructed for stationary long memory with $d > 0$.

If we use the local Whittle estimator, the results for the WLM test get worse and it does not maintain its nominal size - especially for large d. The sup-Wald-LM test, on the other hand, is undersized. However, the other tests reach their nominal size when d is estimated.

When applying the modified local Whittle estimator, we see that the size gets higher for all tests if the sample sizes are small. The WLM test has a liberal size at any value of d. The CUSUM-LM test and the sup-Wald-LM test are also slightly oversized for small d and have higher size distortions if d increases. However, the self-normalized and fixed-b approaches reach their nominal size quickly as the sample size increases.

On the right hand side of the table, we consider additional autoregressive dynamics with $\phi = 0.5$. The size results when assuming a known d get worse for all tests except for the self-normalized and fixed-b approaches. However, if d is estimated using the (modified) local Whittle estimators, all tests except the WLM test are undersized. This can be explained due to the fact that $\widehat{d}_{LW}$ and $\widehat{d}_{mLW}$ are positively biased in finite samples when applied to fractionally integrated processes with additional autoregressive dynamics. Therefore, the critical values for the self-normalized and fixed-b approaches are too high, since they increase with increasing d. For the tests that are based on a consistent long-run variance estimation (WLM, CUSUM-LM, and sup-Wald-LM tests) the critical values decrease as d increases, but the additional scaling in the CUSUM-LM and sup-Wald-LM tests with increasing d balances this effect. For the WLM test this is also true, but the compensation is too slow.

In Table 2 the power results of the introduced tests are given for a fractionally integrated process of type I with one single shift. The shift occurs in the middle of the sample at $\tau = 0.5$ and is as large as the standard deviation of the simulated time series. In general, we observe that the power decreases as d increases, which is due to the fact that means of long-memory processes converge only with a rate of $T^{1/2-d}$ instead of the usual rate of $\sqrt{T}$. This makes the detection of mean shifts in the more persistent time series harder. Furthermore, we see that the power

	$\phi = 0$												$\phi = 0.5$											
	d_0				d_{LW}				d_{mLW}				d_0				d_{LW}				d_{mLW}			
n/d	-0.2	0	0.2	0.4	-0.2	0	0.2	0.4	-0.2	0	0.2	0.4	-0.2	0	0.2	0.4	-0.2	0	0.2	0.4	-0.2	0	0.2	0.4
Fixed-*b* sup-Wald test																								
100	0.043	0.044	0.040	0.051	0.041	0.040	0.040	0.053	0.067	0.069	0.093	0.124	0.077	0.071	0.059	0.067	0.016	0.016	0.014	0.038	0.005	0.009	0.014	0.048
250	0.043	0.044	0.046	0.048	0.042	0.043	0.041	0.048	0.052	0.059	0.065	0.078	0.054	0.049	0.051	0.054	0.018	0.016	0.016	0.029	0.002	0.004	0.007	0.029
500	0.047	0.050	0.046	0.051	0.044	0.044	0.044	0.043	0.050	0.049	0.057	0.068	0.046	0.044	0.044	0.048	0.021	0.018	0.025	0.033	0.003	0.004	0.009	0.028
1000	0.053	0.048	0.045	0.055	0.045	0.048	0.042	0.042	0.050	0.057	0.051	0.058	0.049	0.043	0.047	0.044	0.029	0.023	0.027	0.030	0.005	0.006	0.007	0.028
Self-normalized Wilcoxon test																								
100	0.008	0.045	0.050	0.059	0.008	0.035	0.045	0.062	0.009	0.037	0.076	0.104	0.022	0.070	0.065	0.072	0.013	0.020	0.021	0.048	0.005	0.009	0.020	0.050
250	0.009	0.045	0.050	0.049	0.008	0.038	0.050	0.053	0.009	0.045	0.067	0.077	0.011	0.054	0.059	0.065	0.010	0.023	0.027	0.045	0.003	0.006	0.013	0.043
500	0.010	0.053	0.052	0.060	0.013	0.040	0.053	0.052	0.010	0.043	0.060	0.069	0.009	0.052	0.055	0.057	0.009	0.023	0.031	0.045	0.006	0.008	0.016	0.041
1000	0.010	0.045	0.052	0.060	0.011	0.044	0.050	0.054	0.008	0.052	0.060	0.061	0.010	0.046	0.060	0.052	0.010	0.031	0.036	0.041	0.004	0.009	0.015	0.030
Self-normalized sup-Wald test																								
100	0.043	0.046	0.042	0.050	0.043	0.043	0.042	0.051	0.049	0.053	0.073	0.090	0.065	0.068	0.054	0.059	0.020	0.022	0.018	0.040	0.008	0.011	0.019	0.038
250	0.042	0.046	0.051	0.047	0.042	0.048	0.046	0.051	0.046	0.055	0.061	0.071	0.051	0.054	0.054	0.055	0.025	0.022	0.026	0.035	0.006	0.008	0.012	0.032
500	0.047	0.050	0.049	0.054	0.047	0.047	0.048	0.052	0.045	0.047	0.060	0.063	0.050	0.050	0.050	0.048	0.027	0.026	0.034	0.037	0.008	0.010	0.015	0.028
1000	0.054	0.051	0.046	0.053	0.043	0.050	0.047	0.050	0.052	0.056	0.058	0.056	0.051	0.049	0.056	0.047	0.034	0.031	0.035	0.035	0.011	0.010	0.014	0.032
CUSUM-LM test																								
100	0.029	0.036	0.035	0.038	0.029	0.010	0.007	0.033	0.057	0.072	0.122	0.186	0.105	0.147	0.155	0.149	0.002	0.002	0.010	0.053	0.003	0.007	0.016	0.055
250	0.041	0.046	0.041	0.037	0.029	0.021	0.016	0.046	0.065	0.079	0.111	0.160	0.102	0.115	0.117	0.113	0.003	0.002	0.006	0.034	0.000	0.000	0.002	0.021
500	0.050	0.053	0.047	0.043	0.035	0.025	0.023	0.048	0.072	0.075	0.096	0.131	0.095	0.099	0.102	0.091	0.006	0.004	0.002	0.035	0.000	0.000	0.001	0.009
1000	0.068	0.051	0.049	0.050	0.043	0.032	0.028	0.061	0.073	0.073	0.084	0.112	0.094	0.090	0.088	0.071	0.009	0.005	0.006	0.049	0.000	0.000	0.000	0.005
Sup-Wald-LM test																								
100	0.024	0.036	0.038	0.029	0.010	0.007	0.004	0.001	0.026	0.064	0.122	0.178	0.099	0.159	0.176	0.137	0.002	0.001	0.000	0.000	0.003	0.004	0.006	0.011
250	0.036	0.046	0.047	0.037	0.020	0.015	0.006	0.003	0.051	0.077	0.101	0.144	0.104	0.132	0.120	0.109	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.000
500	0.046	0.050	0.039	0.040	0.032	0.018	0.018	0.004	0.067	0.073	0.098	0.124	0.092	0.113	0.105	0.090	0.005	0.002	0.001	0.000	0.000	0.000	0.000	0.000
1000	0.063	0.055	0.046	0.043	0.041	0.029	0.017	0.006	0.074	0.074	0.086	0.090	0.092	0.099	0.085	0.087	0.011	0.006	0.004	0.000	0.000	0.000	0.000	0.000
Wilcoxon-LM test																								
100	0.137	0.051	0.034	0.106	0.261	0.092	0.052	0.372	0.280	0.153	0.168	0.352	0.746	0.444	0.260	0.309	0.081	0.105	0.495	0.901	0.011	0.216	0.859	0.991
250	0.198	0.057	0.036	0.081	0.283	0.087	0.028	0.316	0.265	0.135	0.128	0.253	0.865	0.495	0.253	0.252	0.181	0.049	0.170	0.875	0.001	0.014	0.779	1.000
500	0.248	0.059	0.037	0.076	0.331	0.079	0.028	0.264	0.289	0.117	0.103	0.181	0.924	0.525	0.259	0.216	0.326	0.099	0.056	0.795	0.001	0.000	0.569	1.000
1000	0.312	0.062	0.037	0.076	0.368	0.076	0.028	0.212	0.333	0.101	0.082	0.145	0.961	0.554	0.264	0.183	0.522	0.170	0.057	0.653	0.000	0.000	0.192	1.000

Table 1: Reported are the size results for fractionally integrated processes ϵ_t of type I with DGP $(1-\phi L)\epsilon_t = (1-L)^{-d}u_t$, bandwidth $m = T^{2/3}$ for $\widehat{d}_{LW}$ and $m = T^{0.8}$, for $\widehat{d}_{mLW}$.

	$\phi = 0$												$\phi = 0.5$											
	d_0				d_{LW}				d_{mLW}				d_0				d_{LW}				d_{mLW}			
n/d	-0.2	0	0.2	0.4	-0.2	0	0.2	0.4	-0.2	0	0.2	0.4	-0.2	0	0.2	0.4	-0.2	0	0.2	0.4	-0.2	0	0.2	0.4
Fixed-b sup-Wald test																								
100	1.000	0.960	0.496	0.202	0.970	0.668	0.287	0.155	1.000	0.932	0.565	0.318	0.967	0.606	0.262	0.153	0.367	0.127	0.086	0.109	0.415	0.133	0.088	0.112
250	1.000	1.000	0.751	0.257	1.000	0.990	0.557	0.203	1.000	1.000	0.785	0.322	1.000	0.929	0.414	0.170	0.983	0.532	0.164	0.119	0.986	0.432	0.123	0.116
500	1.000	1.000	0.918	0.294	1.000	1.000	0.786	0.250	1.000	1.000	0.927	0.348	1.000	0.998	0.576	0.202	1.000	0.932	0.336	0.142	1.000	0.892	0.239	0.142
1000	1.000	1.000	0.989	0.355	1.000	1.000	0.953	0.289	1.000	1.000	0.989	0.383	1.000	1.000	0.794	0.242	1.000	0.999	0.578	0.165	1.000	1.000	0.431	0.167
Self-normalized Wilcoxon test																								
100	0.998	0.948	0.525	0.219	0.961	0.707	0.345	0.181	0.999	0.920	0.560	0.294	0.866	0.639	0.290	0.155	0.512	0.225	0.131	0.121	0.541	0.211	0.141	0.120
250	1.000	1.000	0.742	0.273	1.000	0.974	0.581	0.232	1.000	0.999	0.777	0.310	1.000	0.918	0.447	0.191	0.978	0.635	0.242	0.151	0.983	0.554	0.202	0.137
500	1.000	1.000	0.877	0.305	1.000	1.000	0.760	0.271	1.000	1.000	0.894	0.337	1.000	0.994	0.602	0.217	1.000	0.917	0.409	0.170	1.000	0.890	0.335	0.170
1000	1.000	1.000	0.968	0.359	1.000	1.000	0.903	0.319	1.000	1.000	0.973	0.377	1.000	1.000	0.785	0.254	1.000	0.996	0.601	0.205	1.000	0.995	0.504	0.194
Self-normalized sup-Wald test																								
100	0.997	0.809	0.343	0.134	0.891	0.511	0.204	0.108	0.973	0.801	0.398	0.199	0.997	0.871	0.537	0.312	0.115	0.072	0.085	0.181	0.316	0.138	0.101	0.175
250	1.000	0.992	0.543	0.175	1.000	0.913	0.403	0.137	1.000	0.991	0.593	0.196	1.000	0.997	0.739	0.362	0.821	0.249	0.144	0.181	0.812	0.182	0.064	0.158
500	1.000	1.000	0.736	0.197	1.000	0.995	0.593	0.162	1.000	1.000	0.751	0.226	1.000	1.000	0.878	0.383	1.000	0.803	0.258	0.193	0.998	0.385	0.098	0.156
1000	1.000	1.000	0.882	0.240	1.000	1.000	0.783	0.196	1.000	1.000	0.893	0.251	1.000	1.000	0.974	0.433	1.000	0.999	0.486	0.245	1.000	0.902	0.240	0.135
CUSUM-LM test																								
100	1.000	0.994	0.645	0.235	0.703	0.383	0.175	0.153	0.932	0.929	0.657	0.418	0.997	0.917	0.553	0.392	0.169	0.031	0.063	0.193	0.361	0.121	0.091	0.190
250	1.000	1.000	0.916	0.339	1.000	0.950	0.478	0.258	0.995	0.999	0.883	0.497	1.000	0.999	0.756	0.452	0.888	0.330	0.056	0.185	0.844	0.180	0.038	0.175
500	1.000	1.000	0.993	0.414	1.000	1.000	0.796	0.372	1.000	1.000	0.982	0.566	1.000	1.000	0.886	0.474	1.000	0.879	0.208	0.176	0.999	0.463	0.027	0.177
1000	1.000	1.000	1.000	0.518	1.000	1.000	0.977	0.472	1.000	1.000	1.000	0.619	1.000	1.000	0.977	0.526	1.000	1.000	0.555	0.167	1.000	0.944	0.045	0.158
Sup-Wald-LM test																								
100	1.000	0.986	0.574	0.208	0.390	0.113	0.014	0.001	0.909	0.889	0.590	0.360	0.992	0.825	0.529	0.321	0.017	0.002	0.000	0.000	0.255	0.080	0.036	0.025
250	1.000	1.000	0.886	0.297	0.973	0.776	0.164	0.005	0.992	0.999	0.854	0.412	1.000	0.993	0.709	0.354	0.568	0.076	0.003	0.000	0.711	0.097	0.006	0.001
500	1.000	1.000	0.988	0.384	0.997	0.987	0.493	0.012	1.000	1.000	0.968	0.439	1.000	1.000	0.849	0.371	0.996	0.576	0.036	0.000	0.993	0.266	0.003	0.000
1000	1.000	1.000	1.000	0.472	1.000	1.000	0.871	0.019	1.000	1.000	0.998	0.487	1.000	1.000	0.956	0.426	1.000	0.996	0.237	0.001	1.000	0.828	0.005	0.000
Wilcoxon-LM test																								
100	1.000	0.996	0.651	0.429	0.341	0.408	0.481	0.635	0.999	0.956	0.719	0.636	1.000	0.966	0.661	0.475	0.469	0.542	0.755	0.926	0.539	0.673	0.918	0.994
250	1.000	1.000	0.911	0.482	0.638	0.697	0.586	0.687	1.000	1.000	0.893	0.649	1.000	1.000	0.861	0.524	0.915	0.704	0.680	0.931	0.916	0.619	0.946	1.000
500	1.000	1.000	0.992	0.528	0.925	0.941	0.748	0.728	1.000	1.000	0.980	0.666	1.000	1.000	0.957	0.557	1.000	0.963	0.691	0.907	1.000	0.754	0.942	1.000
1000	1.000	1.000	1.000	0.597	1.000	1.000	0.925	0.740	1.000	1.000	1.000	0.698	1.000	1.000	0.994	0.611	1.000	1.000	0.838	0.866	1.000	0.988	0.910	1.000

Table 2: Reported are the power results for fractionally integrated processes ϵ_t of type I with DGP $\epsilon_t = sd(\epsilon_t)\mathbb{1}(t \geq \lfloor T/2 \rfloor) + (1-L)^{-d}(1-\phi L)^{-1}u_t$, bandwidth $m = T^{2/3}$ for $\widehat{d}_{LW}$ and $m = T^{0.8}$, for $\widehat{d}_{mLW}$.

increases as the sample size increases.

If d is estimated with the local Whittle estimator, the sup-Wald-LM test has substantial power losses for high values of d. As discussed in Section 4.2, this can be explained by the fact that the local Whittle estimate is biased under the alternative hypothesis of a mean shift in the series. Therefore, the critical values are chosen to high. In disregard of the WLM test that shows high size distortions, the FBSW and SNW tests exhibit the highest power among all tests for sample sizes $T < 500$ while the CUSUM-LM tests have the highest power for $T \geq 500$.

If the modified local Whittle estimator is used, the power of the fixed-b and self-normalized approaches increases substantially. This can be explained by the robustness of the modified estimator discussed in Section 4.2. The power gain of the tests based on a consistent estimation of the long run variance (CUSUM-LM, sup-Wald-LM, WLM) is especially pronounced, however, as we observed above, this goes in hand with substantial size distortions.

As in the previous table, at the right hand side of Table 2, we consider additional autoregressive dynamics with $\phi = 0.5$. With the exception of the severely oversized WLM test, all procedures lose power compared to the case with $\phi = 0$. We now observe that the power loss of the CUSUM-LM test is more severe, so that the self-normalized and fixed-b approaches outperform across all sample sizes. The only exception is for large d in small samples, where CUSUM-LM now has the highest power.

Applying the modified local Whittle estimator in this setup does not increase the power. This can be explained by the fact that the estimator captures the short memory dynamics, if a large bandwidth is used. Hou and Perron (2014) suggest using a smaller bandwidth such as $m = \lfloor T^{0.6} \rfloor$ in this situation. Of course in practical applications one does not have the luxury to know whether autoregressive structures are present. Therefore, we keep the bandwidth fixed for all simulations.

Altogether, the results when d is estimated are as follows. Regarding the size results we observe that the self-normalized and fixed-b approaches can be used in applications where d is estimated in any case. The CUSUM-LM and sup-Wald-LM tests should only be used if $\widehat{d}_{LW}$ is applied and the WLM test does not control the size.

The power results show that, when no additional autoregressive component is present, and the local Whittle estimator is used, the FBSW and SNW tests provide the best results in small samples and the CUSUM-LM test in large samples. Furthermore, the power of the self-normalized and fixed-b approaches can be increased significantly if the modified local Whittle estimator is applied. With additional autoregressive components the self-normalized and fixed-b approaches outperform the CUSUM-LM test at any sample size for most values of d.

Our recommendation for practical applications is to use the self-normalized and fixed-b approaches - in particular FBSW and SNW. These tests offer the best size control and also have the highest power in many situations. Whether the local Whittle estimator or its modified version should be applied, depends on the probability that the process exhibits strong autoregressive dynamics. If these are present, the local Whittle estimator should be preferred - otherwise the modified local Whittle estimator of Hou and Perron (2014) offers superior performance in terms of power.

6 Conclusion

In this article, we review recent developments in the area of change-in-mean tests for long-memory time series. The tests are organized according to the testing principles used (CUSUM, sup-Wald, and Wilcoxon) and according to the normalization technique employed (consistent long-run variance estimates or self-normalization).

We then conduct an extensive Monte Carlo simulation study, that specifically addresses the effect of the estimation of the memory parameter. It turns out that this is of particular importance, since the properties of the tests change considerably.

We further find that the performance of the tests is mainly driven by the normalization technique. Altogether, the self-normalized Wilcoxon test of Betken (2016) and the fixed-b sup-Wald test of Iacone, Leybourne, and Taylor (2014) offer the best tradeoff between size control and power.

Finally, the Monte Carlo simulations demonstrate that in absence of strong autoregressive dynamics substantial power gains can be realized if the modified local Whittle estimator of Hou and Perron (2014) is used.

References

Abadir, Karim M., Walter Distaso, and Liudas Giraitis (2009). "Two estimators of the long-run variance: Beyond short memory". In: *Journal of Econometrics* 150(1), pp. 56–70.

Andrews, Donald WK (1991). "Heteroskedasticity and autocorrelation consistent covariance matrix estimation". In: *Econometrica*, pp. 817–858.

Andrews, Donald WK (1993). "Tests for parameter instability and structural change with unknown change point". In: *Econometrica*, pp. 821–856.

Bauer, David F (1972). "Constructing confidence sets using rank statistics". In: *Journal of the American Statistical Association* 67(339), pp. 687–690.

Betken, Annika (2016). "Testing for Change-Points in Long-Range Dependent Time Series by Means of a Self-Normalized Wilcoxon Test". In: *Journal of Time Series Analysis* 37(6), pp. 785–809.

Brown, Robert L, James Durbin, and James M Evans (1975). "Techniques for testing the constancy of regression relationships over time". In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 149–192.

Dehling, Herold, Aeneas Rooch, and Murad S Taqqu (2013). "Non-Parametric Change-Point Tests for Long-Range Dependent Data". In: *Scandinavian Journal of Statistics* 40(1), pp. 153–173.

Diebold, Francis X and Atsushi Inoue (2001). "Long memory and regime switching". In: *Journal of Econometrics* 105(1), pp. 131–159.

Geweke, John and Susan Porter-Hudak (1983). "The estimation and application of long memory time series models". In: *Journal of Time Series Analysis* 4(4), pp. 221–238.

Granger, Clive WJ and Namwon Hyung (2004). "Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns". In: *Journal of Empirical Finance* 11(3), pp. 399–421.

Hidalgo, Javier and Peter M Robinson (1996). "Testing for structural change in a long-memory environment". In: *Journal of Econometrics* 70(1), pp. 159–174.

Horváth, Lajos and Piotr Kokoszka (1997). "The effect of long-range dependence on change-point estimators". In: *Journal of Statistical Planning and Inference* 64(1), pp. 57–81.

Hou, Jie and Pierre Perron (2014). "Modified local Whittle estimator for long memory processes in the presence of low frequency (and other) contaminations". In: *Journal of Econometrics* 182(2), pp. 309–328.

Iacone, Fabrizio, Stephen J Leybourne, and Robert AM Taylor (2014). "A Fixed-B Test for a Break in Level at an Unknown Time Under Fractional Integration". In: *Journal of Time Series Analysis* 35(1), pp. 40–54.

Kiefer, Nicholas M and Timothy J Vogelsang (2005). "A new asymptotic theory for heteroskedasticity-autocorrelation robust tests". In: *Econometric Theory*, pp. 1130–1164.

Krämer, Walter and Philipp Sibbertsen (2002). "Testing for Structural Changes in the Presence of Long Memory". In: *International Journal of Business and Economics* 1(3), pp. 235–242.

Künsch, H. R. (1987). "Statistical aspects of self-similar processes". In: *Proceedings of the First World Congress of the Bernoulli Society* 1, pp. 67–74.

Marinucci, Domenico and Peter M Robinson (1999). "Alternative forms of fractional Brownian motion". In: *Journal of Statistical Planning and Inference* 80(1), pp. 111–122.

McElroy, Tucker and Dimitris N Politis (2012). "Fixed-b asymptotics for the studentized mean from time series with short, long, or negative memory". In: *Econometric Theory* 28(02), pp. 471–481.

Mikosch, Thomas and Cătălin Stărică (2004). "Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects". In: *Review of Economics and Statistics* 86(1), pp. 378–390.

Newey, Whitney K and Kenneth D West (1987). "A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix". In: *Econometrica* 55(3), pp. 703–708.

Ohanissian, Arek, Jeffrey R Russell, and Ruey S Tsay (2008). "True or spurious long memory? A new test". In: *Journal of Business & Economic Statistics* 26(2), pp. 161–175.

Perron, Pierre and Zhongjun Qu (2010). "Long-memory and level shifts in the volatility of stock market return indices". In: *Journal of Business & Economic Statistics* 28(2), pp. 275–290.

Qu, Zhongjun (2011). "A test against spurious long memory". In: *Journal of Business & Economic Statistics* 29(3), pp. 423–438.

Robinson, Peter M (1995a). "Gaussian semiparametric estimation of long range dependence". In: *The Annals of Statistics*, pp. 1630–1661.

Robinson, Peter M (1995b). "Log-periodogram regression of time series with long range dependence". In: *Annals of Statistics* 23(3), pp. 1048–1072.

Robinson, Peter M (2005). "Robust covariance matrix estimation: HAC estimates with long memory/antipersistence correction". In: *Econometric Theory* 21(01), pp. 171–180.

Shao, Xiaofeng (2010). "A self-normalized approach to confidence interval construction in time series". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72(3), pp. 343–366.

Shao, Xiaofeng (2011). "A simple test of changes in mean in the possible presence of long-range dependence". In: *Journal of Time Series Analysis* 32(6), pp. 598–606.

Shao, Xiaofeng and Xianyang Zhang (2010). "Testing for change points in time series". In: *Journal of the American Statistical Association* 105(491), pp. 1228–1240.

Sun, Yixiao (2014). "Let's fix it: Fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference". In: *Journal of Econometrics* 178, pp. 659–677.

Sun, Yixiao, Peter CB Phillips, and Sainan Jin (2008). "Optimal bandwidth selection in heteroskedasticity–autocorrelation robust testing". In: *Econometrica* 76(1), pp. 175–194.

Wang, Lihong (2008). "Change-in-mean problem for long memory time series models with applications". In: *Journal of Statistical Computation and Simulation* 78(7), pp. 653–668.

Wright, Jonathan H (1998). "Testing for a Structural Break at Unknown Date with Long-memory Disturbances". In: *Journal of Time Series Analysis* 19(3), pp. 369–376.

Yajima, Yoshihiro (1988). "On estimation of a regression model with long-memory stationary errors". In: *The Annals of Statistics*, pp. 791–807.