SFB
823

Testing for change in stochastic volatility with long range dependence

Annika Betken, Rafal Kulik

Discussion Paper

Nr. 66/2016

Testing for change in stochastic volatility with long range dependence

Annika Betken [*] Rafał Kulik[†]

November 6, 2016

Abstract

In this paper we consider a change point problem for long memory stochastic volatility models. We show that the limiting behavior for the CUSUM test statistics may not be affected by long memory, unlike the Wilcoxon test statistic which is influenced by long range dependence. We compare our results to subordinated long memory Gaussian processes. Theoretical properties are accompanied by simulation studies.

1 Introduction

One of the most often observed phenomena in financial data is the fact that the log-returns of stock-market prices appear to be uncorrelated, whereas the absolute log-returns or squared log-returns tend to be correlated or even exhibit long-range dependence. Another characteristic of financial time series is the existence of heavy tails in the sense that the marginal tail distribution behaves like a regularly varying function. Both of these features of empirical data can be covered by the so-called long memory stochastic volatility model (LMSV, in short), with its original version introduced in [7]. For this we assume that the data generating process $\{X_j, j \geq 1\}$ satisfies

$$X_j = \sigma(Y_j)\varepsilon_j, \quad j \geq 1 , \tag{1}$$

where

- $\{\varepsilon_j, j \geq 1\}$ is an i.i.d. sequence with mean zero;
- $\{Y_j, j \geq 1\}$ is a stationary, long-range dependent Gaussian process;

[*]Ruhr-Universität Bochum, Fakultät für Mathematik, annika.betken@rub.de; Research supported by the German National Academic Foundation and Collaborative Research Center SFB 823 *Statistical modelling of nonlinear dynamic processes*.

[†]University of Ottawa, Department of Mathematics and Statistics, rkulik@uottawa.ca

- $\sigma(\cdot)$ is a non-negative measurable function, not equal to 0.

Note that within this model long memory results from the subordinated Gaussian sequence $\{Y_j, j \geq 1\}$ only. More precisely, we assume that $\{Y_j, j \geq 1\}$ admits a linear representation with respect to an i.i.d. Gaussian sequence $\{\eta_j, j \geq 1\}$ with $\mathrm{E}(\eta_1) = 0$, $\mathrm{Var}(\eta_1) = 1$, i.e.

$$Y_j = \sum_{k=1}^{\infty} c_k \eta_{j-k} \ , \quad j \geq 0 \ , \tag{2}$$

with $\sum_{k=1}^{\infty} c_k^2 = 1$ and

$$\gamma_Y(k) = \mathrm{Cov}(Y_j, Y_{j+k}) = k^{-D} L_\gamma(k),$$

where $D \in (0,1)$ and L_γ is slowly varying at infinity. Also, we assume that

- $\{(\varepsilon_j, \eta_j), j \geq 1\}$ is a sequence of i.i.d. vectors.

The above set of assumptions we will call collectively **LMSV model**.

The tail behavior of the sequence $\{X_j, j \geq 1\}$ can be related to the tail behavior of ε_j or $\sigma(Y_j)$ or both. Here we will specifically assume that

- random variables ε_j have a marginal distribution with regularly varying right tail, i.e. $\bar{F}_\varepsilon(x) := P(\varepsilon_1 > x) = x^{-\alpha} L(x)$ for some $\alpha > 0$ and a slowly varying function L, such that the following tail balance condition holds:

$$\lim_{x \to \infty} \frac{P(\varepsilon_1 > x)}{P(|\varepsilon_1| > x)} = p = 1 - \lim_{x \to \infty} \frac{P(\varepsilon_1 < -x)}{P(|\varepsilon_1| > x)}$$

  for some $p \in (0,1]$;

- it holds

$$E\left[\sigma^{\alpha+\delta}(Y_0)\right] < \infty$$

  for some $\delta > 0$.

Under these conditions, it follows by Breiman's Lemma (see [17, Proposition 7.5]) that

$$P(X_1 > x) \sim E[\sigma^\alpha(Y_1)] P(\varepsilon_1 > x), \tag{3}$$

i.e. the process $\{X_j, j \geq 1\}$ inherits the tail behavior from $\{\varepsilon_j, j \geq 1\}$.

We would like to point out here that in the literature the usage of the term LMSV often presupposes that the sequences $\{Y_j, j \geq 1\}$ and $\{\varepsilon_j, j \geq 1\}$ are independent. In this paper we will consider a more general model: instead of claiming mutual independence of $\{Y_j, j \geq 1\}$ and $\{\varepsilon_j, j \geq 1\}$, we only assume that (η_j, ε_j) is a sequence of independent random vectors. Especially, this implies that for a fixed index j the random variables ε_j and Y_j are independent while Y_j may depend on $\{\varepsilon_i, i < j\}$. In many cases this version of the LMSV model is referred to as LMSV with leverage.

1.1 Change-point detection under long memory

One of the problems related to financial data is to detect change points in the behavior of the sequence $\{X_j, j \geq 1\}$. Although the problem has been extensively studied for independent random variables (see an excellent book [8]) or in case of weakly dependent data, the issue has not been fully resolved for time series with long memory. The researchers focused rather on justifying whether the observed long range dependence is real or is due to changes in weakly dependent sequences (so-called spurious long memory). See e.g. [3] and Section 7.9.1 of [2] for further references on the latter issue.

As for the testing changes in long memory sequences, one of the first paper seems to be [13], where the authors showed that long range dependence affects the asymptotic behavior of the CUSUM statistics for changes in the mean. For the general testing problem with a change in the marginal distribution under the alternative hypothesis, [12] consider Kolmogorov - Smirnov type change-point tests and change-point estimators for long memory moving average processes. Under the assumption of converging change-point alternatives in LRD time series, the asymptotic behavior of Kolmogorov-Smirnov and Cramér-von Mises type test statistics has also been investigated by [22]. Likewise, [9] show that the Wilcoxon test is always affected by long memory. In fact, in the case of Gaussian long memory data, the asymptotic relative efficiency of the Wilcoxon test and the CUSUM test is 1.

In case of long range dependence, the normalization and the limiting distribution of test statistics typically depend on unknown multiplicative factors or parameters related to the dependence structure of the data generating processes. To bypass estimation of these quantities, the concept of self-normalization has recently been applied to several testing procedures in change-point analysis: In [19] the authors define a self-normalized Kolmogorov-Smirnov test statistic that serves to identify changes in the mean of short range dependent time series. [18] adopted the same approach to define an alternative normalization for the CUSUM test; [4] considers a self-normalized version of the Wilcoxon change-point test proposed by [9].

We refer also to Section 7.9 of [2] for further results on change-point detection for long memory processes.

1.2 Change-point detection for LMSV

In this paper we study CUSUM and Wilcoxon tests for LMSV model and discuss particular cases of testing changes in the mean, in the variance and in the tail index. Of course, the variance and the tail index can be regarded as the mean of transformed random variables, but we observe different effects for each of the three quantities. In particular, the main findings of our paper are as follows:

A-1: CUSUM tests for change in the mean for the LMSV models is typically not affected by long memory (see Corollary 3.2 and Example 4.1). This is different than the findings in [13] for subordinated Gaussian processes;

A-2: Wilcoxon test for change in the mean for the LMSV models is typically affected by long memory (see again Corollary 3.2 and Example 4.1). This is in line with the findings for subordinated Gaussian processes

Hence, to test changes in the mean for the LMSV it is beneficial to use CUSUM test.

B: CUSUM and Wilcoxon tests for change in the variance for the LMSV models are typically affected by long memory; see Section 4.2.

B: CUSUM and Wilcoxon tests for change in the tail index for the LMSV models are typically affected by long memory; see Section 4.3

The paper is structured as follows. In Section 2 we collect some results on subordinated Gaussian processes. In Section 3 we discuss the change-point problem. In particular, we consider CUSUM test and Wilcoxon test. The former is a direct consequence of the existing results, while the latter requires a new theorem on the limiting behavior of empirical processes based on LMSV data. Section 4 is devoted to examples in case of testing changes in the mean, in the variance and in the tail. Since the test statistics and/or limiting distributions involve unknown quantities, self-normalization is considered in Section 5. In fact, in Section 6 we perform simulation studies and indicate that self-normalization provides robustness.

2 Preliminaries

2.1 Some properties of subordinated Gaussian sequences

The main tool for studying the asymptotic behavior of subordinated Gaussian sequences is the Hermite expansion. For a stationary, long-range dependent Gaussian process $\{Y_j, j \geq 1\}$ and a measurable function g such that $E\left(g^2(Y_1)\right) < \infty$ the corresponding Hermite expansion is defined by

$$g(Y_1) - \mathrm{E}\left(g(Y_1)\right) = \sum_{q=m}^{\infty} \frac{J_q(g)}{q!} H_q(Y_1) \ ,$$

where H_q is the q-th Hermite polynomial,

$$J_q(g) = \mathrm{E}\left(g(Y_1) H_q(Y_1)\right)$$

and

$$m = \inf\left\{q \geq 1 \mid J_q(g) \neq 0\right\} \ .$$

The integer m is called the Hermite rank of g and we refer to $J_q(g)$ as the q-th Hermite coefficient of g.

We will also consider the Hermite expansion of the function class

$$\left\{1_{\{g(Y_1)\leq x\}} - F_{g(Y_1)}(x),\ x \in \mathbb{R}\right\} \ ,$$

where $F_{g(Y_1)}$ denotes the distribution function of $g(Y_1)$. For fixed x we have

$$1_{\{g(Y_1)\le x\}} - F_{g(Y_1)}(x) = \sum_{q=m}^{\infty} \frac{J_q(g;x)}{q!} H_q(Y_1)$$

with

$$J_q(g;x) = \mathrm{E}\left(1_{\{g(Y_1)\le x\}} H_q(Y_1)\right) \ .$$

The Hermite rank corresponding to this function class is defined by $m = \inf_x m(x)$, where $m(x)$ denotes the Hermite rank of $1_{\{g(Y_1)\le x\}} - F_{g(Y_1)}(x)$. See [2].

The asymptotic behavior of partial sums of subordinated Gaussian sequences is characterized in [21]. Due to the functional non-central limit theorem in that paper,

$$\frac{1}{d_{n,m}} \sum_{j=1}^{\lfloor nt \rfloor} g(Y_j) \Rightarrow \frac{J_m(g)}{m!} Z_m(t),\ 0 \le t \le 1, \tag{4}$$

where

$$d_{n,m}^2 = \mathrm{Var}\left(\sum_{j=1}^{n} H_m(Y_j)\right) \sim c_m n^{2-mD} L^m(n),\ c_m = \frac{2m!}{(1-Dm)(2-Dm)},$$

$Z_m(t)$, $0 \le t \le 1$, is an m-th order Hermite process and the convergence holds in $\mathbb{D}([0,1])$, the space of functions that are right continuous with left limits. In fact the limiting behavior in (4) is the same as that of the corresponding partial sums based on $\{H_m(Y_j), j \ge 1\}$:

$$J_m(g) \frac{1}{d_{n,m}} \sum_{j=1}^{\lfloor nt \rfloor} H_m(Y_j) \Rightarrow \frac{J_m(g)}{m!} Z_m(t),\ 0 \le t \le 1 \ .$$

Moreover, the functional central limit theorem for the empirical processes was established in [10]. Specifically,

$$\sup_{-\infty \le x \le \infty} \sup_{0 \le t \le 1} d_{n,m}^{-1} \left\{ \lfloor nt \rfloor \left(G_{\lfloor nt \rfloor}(x) - \mathrm{E}\left(G_{\lfloor nt \rfloor}(x)\right)\right) - J_m(g;x) \sum_{j=1}^{\lfloor nt \rfloor} H_m(Y_j) \right\} \xrightarrow{P} 0 \ , \tag{5}$$

where

$$G_l(x) = \frac{1}{l} \sum_{j=1}^{l} 1_{\{g(Y_j)\le x\}}$$

is the empirical distribution function of the sequence $\{g(Y_j), j \ge 1\}$ and $\xrightarrow{P}$ denotes convergence in probability. Thus, the empirical process

$$d_{n,m}^{-1} \lfloor nt \rfloor \left\{ G_{\lfloor nt \rfloor}(x) - \mathrm{E}\left(G_{\lfloor nt \rfloor}(x)\right) \right\} \ , \quad x \in (-\infty, \infty), t \in [0,1] \ ,$$

converges in $\mathbb{D}([-\infty,\infty] \times [0,1])$ to $J_m(g;x) Z_m(t)$.

We refer the reader to [10], [21] and [2] for more details.

3 Change-point problem

Given the observations $X_1, \dots, X_n$ and a function ψ, we define $\xi_j = \psi(X_j)$, $j = 1, \dots, n$, and we consider the testing problem:

$$H_0 : \mathrm{E}(\xi_1) = \dots = \mathrm{E}(\xi_n) ,$$
$$H_1 : \exists\; k \in \{1, \dots, n-1\} \quad \text{such that } \mathrm{E}(\xi_1) = \dots = \mathrm{E}(\xi_k) \neq \mathrm{E}(\xi_{k+1}) = \dots = \mathrm{E}(\xi_n) .$$

We choose ψ according to the specific change-point problem considered. Possible choices include:

- $\psi(x) = x$ in order to detect changes in the mean of the observations $X_1, \dots, X_n$ (change in location);
- $\psi(x) = x^2$ in order to detect changes in the variance of the observations $X_1, \dots, X_n$ (change in volatility);
- $\psi(x) = \log(x^2)$ or $\psi(x) = \log(|x|)$ in order to detect changes in the index α of heavy-tailed observations (change in the tail index).

It is obvious that $\psi(x) = x$ and $\psi(x) = x^2$ lead to testing a change in the mean and variance, respectively. The choice $\psi(x) = \log(|x|)$ requires an additional comment. We note that (3) describes only the asymptotic tail behavior of X_1. For the purpose of this paper, we shall pretend that $P(|X_1| > x) = c^\alpha x^{-\alpha}$, $x > c$, for some $c > 0$. Then the maximum likelihood estimator of $(1/\alpha)$, the reciprocal of the tail index, is

$$\frac{1}{n} \sum_{j=1}^{n} \log(|X_j|/c) \ . \tag{6}$$

This estimator is used in the CUSUM test statistic. To resolve the problem of change-point in the tail index in full generality, we need to employ a completely different technique, based on the so-called tail empirical processes (see [15]). This will be done in a subsequent paper.

In any case, the following test statistics may be applied in order to decide on the change-point problem (H_0, H_1):

- The CUSUM test rejects the hypothesis for large values of the test statistic $C_n = \sup_{0 \le \lambda \le 1} C_n(\lambda)$, where

$$C_n(\lambda) = \left| \sum_{j=1}^{\lfloor n\lambda \rfloor} \psi(X_j) - \frac{\lfloor n\lambda \rfloor}{n} \sum_{j=1}^{n} \psi(X_j) \right| \ . \tag{7}$$

- The Wilcoxon test rejects the hypothesis for large values of the test statistic $W_n = \sup_{0\leq\lambda\leq 1} W_n(\lambda)$, where

$$W_n(\lambda) = \left| \sum_{i=1}^{\lfloor n\lambda \rfloor} \sum_{j=\lfloor n\lambda \rfloor+1}^{n} \left(1_{\{\psi(X_j)\leq\psi(X_j)\}} - \frac{1}{2} \right) \right| . \tag{8}$$

The goal of this paper is to obtain limiting distributions for the CUSUM and the Wilcoxon test statistic in case of time series that follow the LMSV model.

3.1 CUSUM Test for LMSV

In order to determine the asymptotic behavior of the CUSUM test statistic computed with respect to the observations $\psi(X_1), \ldots, \psi(X_n)$, we have to consider the partial sum process $\sum_{j=1}^{\lfloor nt \rfloor} (\psi(X_j) - \mathrm{E}(\psi(X_j)))$.

For the observations $X_1, \ldots, X_n$ that satisfy the LMSV model, the asymptotic behavior of the partial sum process is described by Theorem 4.10 in [2] and hence is stated without the proof. In order to formulate the result, we introduce the following notation:

$$\mathcal{F}_j = \sigma(\varepsilon_j, \varepsilon_{j-1}, \ldots, \eta_j, \eta_{j-1}, \ldots),$$

i.e. $\mathcal{F}_j$ denotes the σ-field generated by the random variables $\varepsilon_j, \varepsilon_{j-1}, \ldots, \eta_j, \eta_{j-1}, \ldots$. Due to the construction ε_j is independent of $\mathcal{F}_{j-1}$ and Y_j is $\mathcal{F}_{j-1}$-measurable.

Theorem 3.1. *Assume that $\{X_j, j \geq 1\}$ follows the LMSV model. Furthermore, assume that $\mathrm{E}(\psi^2(X_1)) < \infty$. Define the function Ψ by $\Psi(y) = \mathrm{E}(\psi(\sigma(y)\varepsilon_1))$. Denote by m the Hermite rank of Ψ and by $J_m(\Psi)$ the corresponding Hermite coefficient.*

1. *If $\mathrm{E}(\psi(X_1) \mid \mathcal{F}_0) \neq 0$ and $mD < 1$, then*

$$\frac{1}{d_{n,m}} \sum_{j=1}^{\lfloor nt \rfloor} (\psi(X_j) - \mathrm{E}(\psi(X_j))) \Rightarrow \frac{J_m(\Psi)}{m!} Z_m(t), \quad t \in [0,1],$$

in $\mathbb{D}([0,1])$.

2. *If $\mathrm{E}(\psi(X_1) \mid \mathcal{F}_0) = 0$, then*

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nt \rfloor} \psi(X_j) \Rightarrow \sigma B(t), \quad t \in [0,1],$$

in $\mathbb{D}([0,1])$, where B denotes a Brownian motion process and $\sigma^2 = \mathrm{E}(\psi^2(X_1))$.

As the immediate consequence of Theorem 3.1 we obtain the asymptotic distribution for the CUSUM statistic.

Corollary 3.2. *Under the assumptions of Theorem 3.1,*

1. *If* $\mathrm{E}(\psi(X_1)\,|\,\mathcal{F}_0) \neq 0$ *and* $mD < 1$*, then*

$$\frac{1}{d_{n,m}} \sup_{0\le\lambda\le 1} C_n(\lambda) \Rightarrow \frac{J_m(\Psi)}{m!} \sup_{0\le t\le 1} \left|Z_m(t) - tZ_m(1)\right| \ . \tag{9}$$

2. *If* $\mathrm{E}(\psi(X_1)\,|\,\mathcal{F}_0) = 0$

$$\frac{1}{\sqrt{n}} \sup_{0\le\lambda\le 1} C_n(\lambda) \Rightarrow \sqrt{\sigma} \sup_{0\le t\le 1} \left|B(t) - tB(1)\right| \ ,$$

where B *denotes a Brownian motion process and* $\sigma^2 = \mathrm{E}(\psi^2(X_1))$.

It is important to note that the Hermite rank of Ψ does not necessarily correspond to the Hermite rank of σ. See Section 4.

3.2 Wilcoxon test for LMSV

For subordinated Gaussian time series $\{g(Y_j), j \ge 1\}$, where $\{Y_j, j \ge 1\}$ is a stationary Gaussian LRD process and g is a measurable function, the asymptotic distribution of the Wilcoxon test statistic W_n is derived from the limiting behavior of the two-parameter empirical process

$$\sum_{j=1}^{\lfloor nt\rfloor} \left(1_{\{g(Y_j)\le x\}} - F_{g(Y_1)}(x)\right) \ , \quad x \in (-\infty, \infty)\ , t \in [0,1] \ ,$$

where $F_{g(Y_1)}$ denotes the distribution function of $g(Y_1)$; see [9].

In order to determine the asymptotic distribution of the Wilcoxon test statistic for the LMSV model, we need to establish an analogous result for the stochastic volatility process $\{X_j, j \ge 1\}$, i.e. our preliminary goal is to prove a limit theorem for the two-parameter empirical process

$$G_n(x,t) = \sum_{j=1}^{\lfloor nt\rfloor} \left(1_{\{\psi(X_j)\le x\}} - F_{\psi(X_1)}(x)\right) ,$$

where now $F_{\psi(X_1)}$ denotes the distribution function of $\psi(X_1)$ with $X_1 = \sigma(Y_1)\varepsilon_1$. To state the weak convergence, we introduce the following notation. Define

$$\Psi_x(y) = P\left(\psi(y\varepsilon_1) \le x\right) \ .$$

Theorem 3.3. *Assume that $\{X_j, j \geq 1\}$ follows the LMSV model. Moreover, assume that*

$$\int_{-\infty}^{\infty} \frac{d}{du} P\left(\psi(u\varepsilon_1) \leq x\right) du < \infty \,. \tag{10}$$

If $mD < 1$, where m denotes the Hermite rank of the class $\left\{1_{\{\sigma(Y_1)\leq x\}} - F_{\sigma(Y_1)}(x),\ x \in \mathbb{R}\right\}$,

$$\frac{1}{d_{n,m}} G_n(x,t) \Rightarrow \frac{J_m(\Psi_x \circ \sigma)}{m!} Z_m(t)\,, \tag{11}$$

in $\mathbb{D}\left([-\infty,\infty] \times [0,1]\right)$.

The proof of this theorem is given in Section 3.3. At this moment we conclude the asymptotic distribution of the Wilcoxon statistics.

Corollary 3.4. *Under the conditions of Theorem 3.3*

$$\frac{1}{n d_{n,m}} \sup_{\lambda\in[0,1]} W_n(\lambda) \Rightarrow \left| \int_{\mathbb{R}} J_m(\Psi_x \circ \sigma) dF_{\psi(X_1)}(x) \right| \frac{1}{m!} \sup_{\lambda\in[0,1]} |Z_m(\lambda) - \lambda Z_m(1)| \,.$$

Proof of Corollary 3.4. According to [9], the asymptotic distribution of the Wilcoxon test statistic can be concluded directly from the limit of the two-parameter empirical process if the sequence $\{X_j, j \geq 1\}$ is ergodic. Ergodicity is obvious since X_j can be represented as a measurable function of the i.i.d. vectors $\{(\eta_j, \varepsilon_j), j \geq 1\}$. □

3.3 Proof of Theorem 3.3

To prove Theorem 3.3, we consider the following decomposition:

$$\begin{aligned} &G_n(x,t) \\ &= \sum_{j=1}^{\lfloor nt \rfloor} \left(1_{\{\psi(X_j)\leq x\}} - \mathrm{E}\left(1_{\{\psi(X_j)\leq x\}} \mid \mathcal{F}_{j-1}\right)\right) + \sum_{j=1}^{\lfloor nt \rfloor} \left(\mathrm{E}\left(1_{\{\psi(X_j)\leq x\}} \mid \mathcal{F}_{j-1}\right) - F_{\psi(X_1)}(x)\right) \\ &=: M_n(x,t) + R_n(x,t). \end{aligned}$$

It will be shown that $n^{-1/2} M_n(x,t) = \mathcal{O}_P(1)$ uniformly in x, t, while $d_{n,m}^{-1} R_n(x,t)$ converges in distribution to the limit process in formula (11). Theorem 3.3 then follows because $\sqrt{n} = \mathrm{o}(d_{n,m})$.

Martingale part. For fixed x the following lemma characterizes the asymptotic behavior of the martingale part $M_n(x,t)$. We write

$$M_n(t) := M_n(x,t) = \sum_{j=1}^{\lfloor nt \rfloor} \zeta_j(x)$$

with $\zeta_j(x) = 1_{\{\psi(X_j)\leq x\}} - \mathrm{E}\left(1_{\{\psi(X_j)\leq x\}} \mid \mathcal{F}_{j-1}\right)$.

Lemma 3.5. *Under the conditions of Theorem 3.3 we have*

$$\frac{1}{\sqrt{n}} M_n(t) \Rightarrow \beta(x) B(t) \ , \quad t \in [0,1] \ ,$$

in $\mathbb{D}([0,1])$*, where* B *denotes a Brownian motion process and* $\beta^2(x) = \mathrm{E}(\zeta_1^2(x))$*.*

Proof. Define

$$\zeta_{nj} = n^{-\frac{1}{2}} \zeta_j(x) = X_{nj}(x) - \mathrm{E}(X_{nj}(x) \mid \mathcal{F}_{j-1})$$

with $X_{nj}(x) = n^{-\frac{1}{2}} 1_{\{\psi(X_j) \leq x\}}$. In order to show convergence in $\mathbb{D}([0,1])$, we apply the functional martingale central limit theorem as stated in Theorem 18.2 of [6]. Therefore, we have to show that

$$\sum_{j=1}^{\lfloor nt \rfloor} \mathrm{E}\left(\zeta_{nj}^2 \mid \mathcal{F}_{j-1}\right) \Rightarrow \beta(x) t$$

for every t and that

$$\lim_{n \to \infty} \sum_{j=1}^{\lfloor nt \rfloor} \mathrm{E}\left(\zeta_{nj}^2 1_{\{|\zeta_{nj}| \geq \epsilon\}}\right) = 0$$

for every t and $\epsilon > 0$ (Lindeberg condition). In order to show that the Lindeberg condition holds, it suffices to show that

$$\lim_{n \to \infty} \sum_{j=1}^{\lfloor nt \rfloor} \mathrm{E}\left(X_{nj}^2(x) 1_{\left\{|X_{nj}(x)| \geq \frac{\epsilon}{2}\right\}}\right) = 0 \qquad (12)$$

due to Lemma 3.3 in [11]. As the indicator function is bounded, the above summands vanish for sufficiently large n and hence (12) follows.

Furthermore, the random variable $\mathrm{E}\left(\zeta_j^2(x) \mid \mathcal{F}_{j-1}\right)$ can be considered as a measurable function of the random variable Y_j and therefore as a function of $\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots$. As a result, $\mathrm{E}\left(\zeta_j^2(x) \mid \mathcal{F}_{j-1}\right)$ is an ergodic sequence and it follows by the ergodic theorem that

$$\frac{1}{n} \sum_{j=1}^{\lfloor nt \rfloor} \mathrm{E}\left(\zeta_j^2(x) \mid \mathcal{F}_{j-1}\right) = \frac{\lfloor nt \rfloor}{n} \frac{1}{\lfloor nt \rfloor} \sum_{j=1}^{\lfloor nt \rfloor} \mathrm{E}\left(\zeta_j^2(x) \mid \mathcal{F}_{j-1}\right) \xrightarrow{P} t \, \mathrm{E}(\zeta_1^2(x))$$

for every t. □

The next lemma establishes tightness of the two-parameter process.

Lemma 3.6. *Under the conditions of Theorem 3.3 we have*

$$\frac{1}{\sqrt{n}} M_n(x,t) = \mathcal{O}_P(1)$$

in $\mathbb{D}([-\infty, \infty] \times [0,1])$*.*

The (technical) proof of this lemma is postponed to Section 7.

Long memory part. Finally, we prove weak convergence of the long memory part $R_n(x,t)$.

Lemma 3.7. *Under the conditions of Theorem 3.3,*

$$\frac{1}{d_{n,m}} R_n(x,t) \Rightarrow \frac{J_m(\Psi_x \circ \sigma)}{m!} Z_m(t) \ ,$$

in $\mathbb{D}\left([-\infty,\infty]\times[0,1]\right)$.

Proof. Note that

$$\mathrm{E}\left(1_{\{\psi(X_j)\le x\}} \mid \mathcal{F}_{j-1}\right) = \mathrm{E}\left(1_{\{\psi(\sigma(Y_j)\varepsilon_j)\le x\}} \mid \mathcal{F}_{j-1}\right) = \Psi_x(\sigma(Y_j))$$

because Y_j is $\mathcal{F}_{j-1}$-measurable and ε_j is independent of $\mathcal{F}_{j-1}$. Furthermore, $\mathrm{E}\left(\Psi_x(\sigma(Y_j))\right) = F_{\psi(X_1)}(x)$, the distribution function of $\psi(X_1)=\psi(\sigma(Y_1)\varepsilon_1)$. Hence,

$$\begin{aligned} R_n(x,t) &= \sum_{j=1}^{\lfloor nt\rfloor} \left(\Psi_x(\sigma(Y_j)) - F(x)\right) \\ &= \lfloor nt\rfloor \int_{-\infty}^{\infty} \Psi_x(u) d\left(G_{\lfloor nt\rfloor} - \mathrm{E}\, G_{\lfloor nt\rfloor}\right)(u), \end{aligned}$$

where

$$G_l(u) = \frac{1}{l}\sum_{j=1}^{l} 1_{\{\sigma(Y_j)\le u\}}$$

is the empirical distribution function of the sequence $\{\sigma(Y_j), j\ge 1\}$. We have,

$$\begin{aligned} & d_{n,m}^{-1} R_n(x,t) \\ &= d_{n,m}^{-1}\lfloor nt\rfloor \int_{-\infty}^{\infty} \Psi_x(u) d\left(G_{\lfloor nt\rfloor} - \mathrm{E}\, G_{\lfloor nt\rfloor}\right)(u) \\ &= -\left\{\int_{-\infty}^{\infty} \frac{d}{du} P\left(\psi(u\varepsilon_1)\le x\right) d_{n,m}^{-1} \left\{\lfloor nt\rfloor \left[G_{\lfloor nt\rfloor}(u) - \mathrm{E}\, G_{\lfloor nt\rfloor}(u)\right] - \frac{J_m(\sigma;u)}{m!}\sum_{j=1}^{\lfloor nt\rfloor} H_m(Y_j)\right\} du\right\} \\ &\quad - \left\{\int_{-\infty}^{\infty} \frac{d}{du} P\left(\psi(u\varepsilon_1)\le x\right) d_{n,m}^{-1} \frac{J_m(\sigma;u)}{m!}\sum_{j=1}^{\lfloor nt\rfloor} H_m(Y_j) du\right\} =: I_1(x,t) + I_2(x,t), \end{aligned}$$

where m denotes the Hermite rank of the class $\left\{1_{\{\sigma(Y_1)\le x\}} - F_{\sigma(Y_1)}(x),\ x\in\mathbb{R}\right\}$ and

$$J_m(\sigma;y) = \mathrm{E}\left(1_{\{\sigma(Y_1)\le y\}} H_m(Y_1)\right) \ .$$

Using the reduction principle (5) with $g = \sigma$ and the integrability condition (10), we conclude that the first summand converges to 0 in probability, uniformly in x, t. Furthermore,

$$\begin{aligned} I_2(x,t) &= -\left\{ \int_{-\infty}^{\infty} \frac{d}{du} P\left(\psi(u\varepsilon_1) \le x\right) d_{n,m}^{-1} \frac{J_m(\sigma;u)}{m!} \sum_{j=1}^{\lfloor nt \rfloor} H_m(Y_j) du \right\} \\ &= -d_{n,m}^{-1} \sum_{j=1}^{\lfloor nt \rfloor} H_m(Y_j) \left\{ \int_{-\infty}^{\infty} \frac{d}{du} P\left(\psi(u\varepsilon_1) \le x\right) \frac{J_m(\sigma;u)}{m!} du \right\}. \end{aligned}$$

We have

$$d_{n,m}^{-1} \sum_{j=1}^{\lfloor nt \rfloor} H_m(Y_j) \Rightarrow Z_m(t).$$

Moreover, integration by parts yields

$$\begin{aligned} &\int_{-\infty}^{\infty} \frac{d}{du} P\left(\psi(u\varepsilon_1) \le x\right) J_m(\sigma;u) du \\ &= \int_{-\infty}^{\infty} \frac{d}{du} P\left(\psi(u\varepsilon_1) \le x\right) \int 1_{\{\sigma(z) \le u\}} H_m(z) \varphi(z) dz du \\ &= \int_{-\infty}^{\infty} H_m(z) \varphi(z) \int \frac{d}{du} P\left(\psi(u\varepsilon_1) \le x\right) 1_{\{\sigma(z) \le u\}} du dz \\ &= \int_{-\infty}^{\infty} H_m(z) \varphi(z) \int_{\sigma(z)}^{\infty} \frac{d}{du} P\left(\psi(u\varepsilon_1) \le x\right) du dz \\ &= \lim_{u \to \infty} P\left(\psi(u\varepsilon_1) \le x\right) \int_{-\infty}^{\infty} H_m(z) \varphi(z) dz - \int_{-\infty}^{\infty} P(\psi(\sigma(z)\varepsilon_1) \le x) H_m(z) \varphi(z) dz \\ &= -J_m(\Psi_x \circ \sigma). \end{aligned}$$

□

4 Examples

4.1 Change in the mean

To test a change in the mean we choose $\psi(x) = x$.

CUSUM: Recall that the function Ψ in Theorem 3.1 is defined as $\Psi(y) = \mathrm{E}\left(\psi(\sigma(y)\varepsilon_1)\right)$. In this case

$$\mathrm{E}(\psi(X_1) \mid \mathcal{F}_0) = \sigma(Y_1)\, \mathrm{E}(\varepsilon_1) = 0.$$

Therefore, the CUSUM statistic converges to a Brownian bridge. Hence,

- Long memory does not influence the asymptotic behavior of the CUSUM statistic for testing change in the mean.

Wilcoxon: Recall that $\Psi_x(y) = P\left(\psi(y\varepsilon_1) \le x\right)$. Using the integration by parts and noting that $(d/dz)\varphi(z) = -z\varphi(z)$ we have

$$\begin{aligned}
J_1(\Psi_x \circ \sigma) &= \int_{-\infty}^{\infty} P\left(\psi(\sigma(z)\varepsilon_1) \le x\right) z\varphi(z)dz \\
&= \int_{-\infty}^{\infty} \frac{d}{dz} P\left(\psi(\sigma(z)\varepsilon_1) \le x\right) \varphi(z)dz = \int_{-\infty}^{\infty} \frac{d}{dz} P\left(\varepsilon_1 \le \frac{x}{\sigma(z)}\right) \varphi(z)dz \\
&= \int_{-\infty}^{\infty} \left(\frac{1}{\sigma(z)}\right)' f_\varepsilon\left(\frac{x}{\sigma(z)}\right) \varphi(z)dz \,,
\end{aligned}$$

where f_ε is the density of ε_1 (if it exists). Here, different scenarios are possible. If $\sigma(y) = y^2$ then $z \to \left(\frac{1}{\sigma(z)}\right) f_\varepsilon(x/\sigma(z))'\varphi(z)$ is antisymmetric for any x and any choice of f_ε. Hence, $J_1(\Psi_x \circ \sigma) = 0$ and one can calculate that the Hermite rank of $\Psi_x \circ \sigma$ is 2. If $\sigma(y) = \exp(y)$ and e.g. ε_1 is Pareto-distributed, i.e. for some $\alpha > 0$, $c > 0$

$$f_\varepsilon(x) = \begin{cases} \frac{\alpha c^\alpha}{x^{\alpha+1}}, & x \ge c \\ 0, & x < c \,, \end{cases} .$$

then, as a result,

$$\begin{aligned}
&\int_{-\infty}^{\infty} e^{-z} \frac{\alpha c^\alpha}{\left(\frac{x}{\exp(z)}\right)^{\alpha+1}} 1_{\left\{\frac{x}{\exp(z)} \ge c\right\}} \varphi(z)dz \\
&= \alpha c^\alpha x^{-(\alpha+1)} \int_{-\infty}^{\infty} \exp(z\alpha) 1_{\left\{\log\left(\frac{x}{c}\right) \ge z\right\}} \varphi(z)dz \\
&= \alpha c^\alpha x^{-(\alpha+1)} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\log\left(\frac{x}{c}\right)} \underbrace{\exp(z\alpha - \frac{1}{2}z^2)}_{>0} dz \,.
\end{aligned}$$

Hence, $J_1(\Psi_x \circ \sigma) \neq 0$. In any case,

- Long memory influences the asymptotic behavior of the Wilcoxon statistic, unlike the CUSUM one.

4.2 Change in the variance

To test a change in the variance we choose $\psi(x) = x^2$.

CUSUM: Recall again that the function Ψ in Theorem 3.1 is defined as $\Psi(y) = \mathrm{E}\left(\psi(\sigma(y)\varepsilon_1)\right)$. Then

$$\mathrm{E}(\psi(X_1) \mid \mathcal{F}_0) = \sigma^2(Y_1)\, \mathrm{E}(\varepsilon_1^2) \neq 0$$

and hence long memory affects the limiting behavior of the CUSUM statistic. Moreover,

$$J_m(\Psi) = \mathrm{E}(\varepsilon_1^2) \int \sigma^2(z) H_m(z) \varphi(z) dz = \mathrm{E}\left(\varepsilon_1^2\right) J_m(\sigma^2),$$

i.e. the Hermite rank of Ψ equals the Hermite rank of σ^2. If $mD < 1$ then the limiting behavior of the CUSUM statistic is described by (9). Hence,

- Long memory influences the asymptotic behavior of the CUSUM statistic for testing change in the variance.

Wilcoxon: Recall again that $\Psi_x(y) = P\left(\psi(y\varepsilon_1) \leq x\right)$. We have

$$\begin{aligned} J_1(\Psi_x \circ \sigma) &= \int_{-\infty}^{\infty} P\left(\psi(\sigma(z)\varepsilon_1) \leq x\right) z \varphi(z) dz \\ &= \int_{-\infty}^{\infty} \frac{d}{dz} P\left(\psi(\sigma(z)\varepsilon_1) \leq x\right) \varphi(z) dz = \int_{-\infty}^{\infty} \frac{d}{dz} P\left(\varepsilon_1^2 \leq \frac{x}{\sigma^2(z)}\right) \varphi(z) dz \\ &= \int_{-\infty}^{\infty} \left(\frac{1}{\sigma^2(z)}\right)' f_\varepsilon \left(\frac{x}{\sigma^2(z)}\right) \varphi(z) dz \;. \end{aligned}$$

If $\sigma(y) = \exp(-y)$ then we are in the same situation as in case of testing the mean and hence $J_1(\Psi_x \circ \sigma) \neq 0$. Hence,

- Long memory influences the asymptotic behavior of the Wilcoxon statistic for testing change in the variance.

4.3 Change in the tail index

To test a change in the tail index we choose $\psi(x) = \log(x^2)$.

CUSUM: In this case

$$\mathrm{E}(\psi(X_1) \mid \mathcal{F}_0) = \log(\sigma^2(Y_1)) + \mathrm{E}(\log(\varepsilon_1^2)) \neq 0$$

and hence long memory affects the limiting distribution of the CUSUM statistic. Moreover,

$$J_m(\Psi) = 2 \int \log(\sigma(z)) H_m(z) \varphi(z) dz = 2 J_m(\log \circ \sigma),$$

so that the Hermite rank of Ψ equals the Hermite-rank of $h = \log \circ \sigma$.

We note further that in case of $\psi(x) = \log(x^2)$ we have

$$\begin{aligned} &\frac{1}{d_{n,m}} \sum_{j=1}^{\lfloor n\lambda \rfloor} \left(\log\left(X_j^2\right) - \mathrm{E}\log\left(X_j^2\right)\right) \\ &= \frac{2}{d_{n,m}} \sum_{j=1}^{\lfloor n\lambda \rfloor} \left(\log \sigma(Y_j) - \mathrm{E}\log \sigma(Y_j)\right) + \frac{\sqrt{n}}{d_{n,m}} \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor n\lambda \rfloor} \left(\log\left(\varepsilon_j^2\right) - \mathrm{E}\log\left(\varepsilon_j^2\right)\right). \end{aligned}$$

The first summand converges to $(J_m(\log \circ\sigma)/m!)Z_m(\lambda)$ as a consequence of the functional non-central limit theorem discussed in Section 2.1. The second term is $o_P(1)$ uniformly in λ by Donsker's theorem, if $\mathrm{Var}(\log \varepsilon_1^2) < \infty$. As a result, this observations is consistent with Corollary 3.2.

In summary

- Long memory influences the asymptotic behavior of the CUSUM statistic for testing change in the tail index.

5 Self-normalization

An application of the CUSUM test presupposes knowledge of the normalizing sequence $d_{n,m}$ (if $\mathrm{E}(\psi(X_1) \mid \mathcal{F}_0) \neq 0$) and of the coefficients $J_m(\Psi)$ or σ that appear in the limit of the test statistic. Usually, these quantities are unknown. In order to avoid estimation of the normalization and the unknown coefficients in the limit, we consider the self-normalized CUSUM test statistic with respect to the observations $\xi_j = \psi(X_j)$, $j = 1, \dots, n$. For $0 < \tau_1 < \tau_2 < 1$ it is defined by

$$T_n(\tau_1, \tau_2) = \sup_{k \in \{\lfloor n\tau_1 \rfloor, \dots, \lfloor n\tau_2 \rfloor\}} |G_n(k)|,$$

where

$$G_n(k) = \frac{\sum_{j=1}^{k} \xi_j - \frac{k}{n} \sum_{j=1}^{n} \xi_j}{\left\{\frac{1}{n} \sum_{t=1}^{k} S_t^2(1,k) + \frac{1}{n} \sum_{t=k+1}^{n} S_t^2(k+1,n)\right\}^{\frac{1}{2}}},$$

with

$$S_t(j,k) = \sum_{h=j}^{t} \left(\xi_h - \bar{Z}_{j,k}\right),$$

$$\bar{Z}_{j,k} = \frac{1}{k-j+1} \sum_{t=j}^{k} \xi_t.$$

The self-normalized CUSUM test rejects the hypothesis for large values of the test statistic $T_n(\tau_1, \tau_2)$. Note that the proportion of the data that is included in the calculation of the supremum is restricted by τ_1 and τ_2. A common choice is $\tau_1 = 1 - \tau_2 = 0.15$; see [1].

In order to detect changes in the mean of (possibly) long-range dependent time series, a similar test statistic has been proposed by in [18]. For long memory stochastic volatility sequences the limit of the test statistic can be derived in the same way as in [18]. Under the assumptions of Theorem 3.1 (and if $\mathrm{E}(\psi(X_1) \mid \mathcal{F}_0) \neq 0$ and $mD < 1$), an application of the continuous mapping theorem to the partial sum process $\frac{1}{d_n}\sum_{j=1}^{\lfloor nt\rfloor}\left(\psi(X_j)-\operatorname{E}\psi(X_j)\right)$ yields $T_n(\tau_1,\tau_2) \xrightarrow{\mathcal{D}} T(m,\tau_1,\tau_2)$, where

$$T(m,\tau_1,\tau_2)=\sup_{\lambda\in[\tau_1,\tau_2]}\frac{|Z_m(\lambda)-\lambda Z_m(1)|}{\left\{\int_0^\lambda V_m^2(r;0,\lambda)dr+\int_\lambda^1 V_m^2(r;\lambda,1)dr\right\}^{\frac{1}{2}}}$$

with

$$V_m(r;r_1,r_2)=Z_m(r)-Z_m(r_1)-\frac{r-r_1}{r_2-r_1}\left\{Z_m(r_2)-Z_m(r_1)\right\}$$

for $r\in[r_1,r_2]$, $0<r_1<r_2<1$.

If $\operatorname{E}(\psi(X_1)\,|\,\mathcal{F}_0)=0$, it follows by an application of the continuous mapping theorem to the partial sum process $\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor nt\rfloor}\psi(X_j)$ that $T_n(\tau_1,\tau_2) \xrightarrow{\mathcal{D}} T(\tau_1,\tau_2)$, where

$$T(\tau_1,\tau_2)=\sup_{\lambda\in[\tau_1,\tau_2]}\frac{|B(\lambda)-\lambda B(1)|}{\left\{\int_0^\lambda V^2(r;0,\lambda)dr+\int_\lambda^1 V^2(r;\lambda,1)dr\right\}^{\frac{1}{2}}}$$

with

$$V(r;r_1,r_2)=B(r)-B(r_1)-\frac{r-r_1}{r_2-r_1}\left\{B(r_2)-B(r_1)\right\}$$

for $r\in[r_1,r_2]$, $0<r_1<r_2<1$. Note that in this case the limit does not depend on any unknown parameters: for one thing the factor σ that appeared in the limit of the partial sum process is canceled out by self-normalization, for another thing the limit does neither depend on the Hermite rank m nor on the Hurst parameter H.

6 Simulations

6.1 Change in the mean

We will now investigate the finite sample performance of the CUSUM and Wilcoxon change-point test for detecting changes in the mean of LMSV time series $\{X_j, j\geq 1\}$, i.e. we choose $\psi(x)=x$ for the test statistics C_n and W_n as described in Section 3.

For the simulations we make the following specifications:

$$X_j=\sigma(Y_j)\varepsilon_j,\quad j\geq 0\ , \tag{13}$$

where

- $\{\varepsilon_j, j\geq 1\}$ are standard normal generated by the function `rnorm` in `R`;

- $\{Y_j, j \geq 1\}$ is a fractional Gaussian noise generated by the function `fgnSim` (`fArma` package in `R`) with the Hurst parameter H (note that H and the memory parameter D are linked by $H = 1 - D/2$);

- $\sigma(y) = \exp(y)$.

In order to determine the finite sample performance under the alternative, we simulate time series with a change-point of height h after τ percent of the data, i.e. we consider random variables X_j, $j = 1, \ldots, n$ with expected value μ_j, $j = 1 \ldots, n$ such that $\mu_j = \mu$ for $j = 1, \ldots, \lfloor n\tau \rfloor$ while $\mu_j = \mu + h$ for $j = \lfloor n\tau \rfloor + 1, \ldots, n$.

Recall that the function Ψ in Theorem 3.1 is defined as $\Psi(y) = \mathrm{E}\left(\psi(\sigma(y)\varepsilon_1)\right)$. In this case

$$\mathrm{E}(\psi(X_1) \mid \mathcal{F}_0) = \sigma(Y_1)\,\mathrm{E}(\varepsilon_1) = 0.$$

Due to Corollary 3.2

$$\frac{1}{\sqrt{n}} \sup_{0 \leq \lambda \leq 1} C_n(\lambda) \Rightarrow \sqrt{\sigma} \sup_{0 \leq t \leq 1} \left|B(t) - tB(1)\right| \ ,$$

where $\sigma^2 = \mathrm{E}(\psi^2(X_1)) = \mathrm{E}(\varepsilon_1^2 \exp(2Y_1)) = \mathrm{E}(\exp(2Y_1)) = \exp(2)$. Hence, we expect long memory in the data not to influence the asymptotic behavior of the CUSUM statistic for testing change in the mean. In fact, our simulations confirm that the power of the CUSUM point test does not change significantly when different values for H are considered; see Table 1.

In order to compute the asymptotic distribution of the Wilcoxon test in the above situation, note that

$$\begin{aligned} J_1(\Psi_x \circ \sigma) &= \int_{-\infty}^{\infty} \left(\frac{1}{\sigma(z)}\right)' f_{\varepsilon}\left(\frac{x}{\sigma(z)}\right) \varphi(z) dz \\ &= -\int_{-\infty}^{\infty} \mathrm{e}^{-z}\, \varphi\left(x\, \mathrm{e}^{-z}\right)\right) \varphi(z) dz \\ &= -\int_{0}^{\infty} \varphi(xz) \varphi\left(\log(z)\right) dz \ . \end{aligned}$$

Obviously, the expression on the right-hand side does not equal 0 for any $x \in \mathbb{R}$. Therefore, we have

$$\frac{1}{n d_{n,1}} \sup_{\lambda \in [0,1]} W_n(\lambda) \Rightarrow \left| \int_{\mathbb{R}} J_1(\Psi_x \circ \sigma) dF_{X_1}(x) \right| \sup_{\lambda \in [0,1]} \left|B_H(\lambda) - \lambda B_H(1)\right| \ ,$$

where $d_{n,1} \sim \sqrt{c_1} n^{1-\frac{D}{2}} L^{\frac{1}{2}}(n)$, $c_1 = 2((1-D)(2-D))^{-1}$. For fractional Gaussian noise, $L(n) \sim \frac{(1-D)(2-D)}{2}$ so that $d_{n,1} \sim n^{1-\frac{D}{2}}$. In order to determine the multiplicative factor

	n	H	**Wilcoxon** $h=0.5$	$h=1$	$h=2$	**CUSUM** $h=0.5$	$h=1$	$h=2$
$\tau=0.25$	500	0.6	0.5548	0.9984	1.0000	0.0000	0.0002	0.0968
	1000		0.9334	1.0000	1.0000	0.0002	0.0020	0.9536
	2000		1.0000	1.0000	1.0000	0.0000	0.1032	1.0000
	500	0.7	0.0386	0.8036	1.0000	0.0002	0.0006	0.1022
	1000		0.1438	0.9896	1.0000	0.0000	0.0018	0.9520
	2000		0.5062	1.0000	1.0000	0.0002	0.1100	1.0000
	500	0.8	0.0012	0.1854	0.8770	0.0004	0.0010	0.0930
	1000		0.0024	0.3614	0.9804	0.0004	0.0054	0.9544
	2000		0.0058	0.6284	0.9988	0.0004	0.1026	0.9998
	500	0.9	0.0078	0.1152	0.5234	0.0030	0.0052	0.0934
	1000		0.0076	0.1454	0.6134	0.0038	0.0126	0.9638
	2000		0.0086	0.1764	0.7028	0.0052	0.0928	0.9996
$\tau=0.5$	500	0.6	0.9066	1.0000	1.0000	0.0000	0.0012	0.8324
	1000		0.9982	1.0000	1.0000	0.0000	0.0634	1.0000
	2000		1.0000	1.0000	1.0000	0.0012	0.8316	1.0000
	500	0.7	0.3090	0.9908	1.0000	0.0002	0.0028	0.8362
	1000		0.6860	0.9998	1.0000	0.0000	0.0780	1.0000
	2000		0.9764	1.0000	1.0000	0.0024	0.8244	1.0000
	500	0.8	0.0378	0.6398	0.9940	0.0008	0.0052	0.8504
	1000		0.0740	0.8538	0.9998	0.0008	0.0672	0.9996
	2000		0.1482	0.9696	1.0000	0.0034	0.8448	1.0000
	500	0.9	0.0552	0.3738	0.8250	0.0052	0.0118	0.8812
	1000		0.0646	0.4186	0.8754	0.0052	0.0644	0.9990
	2000		0.0656	0.4990	0.9306	0.0096	0.8768	1.0000

Table 1: Rejection rates of the CUSUM and Wilcoxon change-point test for stochastic volatility time series of length n which satisfy (13). The calculations are based on 5,000 simulation runs.

that appears in the limit of the Wilcoxon test statistic, we have to determine the density f_X of X_1. It holds that

$$f_X(x) = \int \frac{1}{|t|} f_{\sigma(Y_1)}(t) f_\varepsilon\left(\frac{x}{t}\right) dt = \int \frac{1}{t^2} \varphi(\log(t)) \varphi\left(\frac{x}{t}\right) dt = \int_0^\infty \varphi(xt) \varphi\left(\log(t)\right) dt \ .$$

As a result, we get

$$\left| \int_{\mathbb{R}} J_1(\Psi_x \circ \sigma) dF_{X_1}(x) \right| = \left| \int_{\mathbb{R}} G^2(x) dx \right| \ ,$$

where

$$G(x) = \int_0^\infty \varphi(xt) \varphi\left(\log(t)\right) dt \ .$$

Numerical integration yields $\left|\int_{\mathbb{R}} J_1(\Psi_x \circ \sigma) dF_{X_1}(x)\right| \approx 0.2765841$. Clearly, the asymptotic behavior of the Wilcoxon statistic is influenced by the intensity of dependence in the data since the limit distribution depends on the Hurst parameter H. This observation is reflected in our simulation results in Table 1: An increase of dependence in the data goes along with a significant loss of power. A natural explanation for this phenomenon is given by the fact that a growth of dependence (in H) in time series entails an increase of the variance, so that it becomes harder to detect a level shift of a fixed height.

6.2 Change in the tail

We will now investigate the finite sample performance of the CUSUM change-point test for detecting changes in the tail parameter α of LMSV time series $\{X_j, j \geq 1\}$, i.e. we choose $\psi(x) = \log(|x|)$ for the test statistic C_n as described in Section 3. In this case the Hermite rank of Ψ equals $m = 1$, $J_m(\Psi) = 1$ and the normalization is $d_{n,1}$.

For the simulations we make the following specifications:

$$X_j = \sigma(Y_j)\varepsilon_j, \quad j \geq 0 \ , \tag{14}$$

where

- $\{\varepsilon_j, j \geq 1\}$ are Pareto distributed with parameter α generated by the function `rgpd` (`fExtremes` package in `R`);
- $\{Y_j, j \geq 1\}$ is fractional Gaussian noise generated by the function `fgnSim` (`fArma` package in `R`) with the Hurst parameter H (note that H and the memory parameter D are linked by $H = 1 - D/2$);
- $\sigma(y) = \exp(y)$.

In order to determine the finite sample performance under the alternative, we simulate time series with a change-point of height h after τ percent of the data, i.e. we consider Pareto distributed random variables ε_j, $j = 1, \ldots, n$ with parameters α_j, $j = 1 \ldots, n$ such that $\alpha_j = \alpha$ for $j = 1, \ldots, \lfloor n\tau \rfloor$ while $\alpha_j = \alpha + h$ for $j = \lfloor n\tau \rfloor + 1, \ldots, n$.

The simulation results in Table 2 show that, under the null hypothesis, the rejection rates of both testing procedures (the classical CUSUM and its self-normalized version) approach the significance level of 5% whenever α is not extremely small. If $\alpha = 0.25$ or $\alpha = 0.05$, the CUSUM test performs badly, while the self-normalized CUSUM works well. This is due to the fact that extremely large observations (due to small values of α) are misinterpreted as change-points. Furthermore, we can see from Table 3 that both tests detect change-points that are located in the middle of the sample with higher frequency than change-points that are located close to the boundary of the testing region. In the presence of a change in α, the number of test decisions in favor of the alternative increases as the height h of the level shift in the tail index increases. Moreover, it is notable that small values of α go along with a high number of rejections under the alternative as well. In particular, one can observe that the heavier the tails (i.e. the smaller α) the higher the rejection rate. This is very likely due to the aforementioned misinterpretation of the extremely large values as changes.

Comparing the rejection rates of the tests for different Hurst parameters, the following relation seems to hold: the bigger H, the smaller the number of rejections under the hypothesis and under the alternative. Under the alternative this may be due to the fact that the variance of the observations increases as the dependence, i.e. the values of the parameter H, increases (as suggested in Section 4.3). Therefore, it becomes harder to detect a level shift of a fixed height. The difference in rejection rates for different H is less prominent in the self-normalized version of the test.

In general:

- The self-normalized test seems to be too conservative in the sense that the empirical size of the testing procedure is smaller than the significance level;
- The rejection rate of the classical CUSUM test is considerably higher than the nominal level of 5%.
- Under the alternative, the rejection rate of the classical CUSUM test is bigger than that of the self-normalized test for any choice of the parameters α and H.

As for the latter comment, in the context of change-point tests the so-called "better size but less power" phenomenon based on self-normalized statistics has been observed by [18], [19] and [4]. Moreover, it is consistent with observations made in [16] and [20]. It is important to note that the finite sample results are based on simulations which were executed under the assumption that the normalization $d_{n,m}$, which depends on the parameters m, D and the slowly varying function L, and the coefficient $J_m(\Psi)$ are known. For all practical purposes this is not the case, so that both expressions have to be estimated. In contrast, the self-normalized test statistic can be computed from the given data while its limit depends on the parameters m and D only. For an adequate comparison of the testing procedures this has to be taken into consideration.

		CUSUM				**self-normalized CUSUM**			
n	H	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$
50	0.6	0.961	0.567	0.168	0.065	0.024	0.031	0.032	0.041
	0.7	0.902	0.428	0.126	0.060	0.012	0.019	0.030	0.045
	0.8	0.849	0.343	0.115	0.060	0.005	0.010	0.024	0.042
	0.9	0.847	0.340	0.099	0.058	0.005	0.004	0.021	0.039
100	0.6	0.971	0.544	0.156	0.066	0.022	0.031	0.037	0.045
	0.7	0.880	0.366	0.107	0.060	0.019	0.024	0.035	0.047
	0.8	0.754	0.248	0.097	0.060	0.007	0.015	0.030	0.043
	0.9	0.701	0.223	0.083	0.057	0.005	0.009	0.028	0.042
200	0.6	0.969	0.510	0.148	0.068	0.024	0.029	0.038	0.047
	0.7	0.819	0.296	0.096	0.060	0.016	0.026	0.041	0.052
	0.8	0.605	0.191	0.074	0.056	0.008	0.020	0.037	0.042
	0.9	0.492	0.146	0.075	0.053	0.007	0.021	0.044	0.047
300	0.6	0.964	0.486	0.146	0.067	0.0267	0.035	0.040	0.051
	0.7	0.778	0.262	0.095	0.056	0.019	0.026	0.042	0.047
	0.8	0.506	0.161	0.075	0.057	0.008	0.023	0.038	0.044
	0.9	0.380	0.112	0.056	0.042	0.009	0.025	0.038	0.042
400	0.6	0.958	0.473	0.139	0.069	0.029	0.033	0.041	0.048
	0.7	0.729	0.249	0.093	0.063	0.016	0.031	0.046	0.050
	0.8	0.461	0.135	0.062	0.048	0.013	0.026	0.034	0.046
	0.9	0.312	0.100	0.059	0.046	0.010	0.027	0.040	0.052
500	0.6	0.955	0.454	0.138	0.065	0.025	0.030	0.043	0.046
	0.7	0.701	0.222	0.087	0.056	0.019	0.030	0.047	0.049
	0.8	0.402	0.126	0.065	0.058	0.013	0.027	0.041	0.046
	0.9	0.263	0.085	0.057	0.049	0.014	0.029	0.044	0.052
1000	0.6	0.937	0.424	0.125	0.065	0.032	0.037	0.041	0.047
	0.7	0.595	0.179	0.073	0.056	0.022	0.032	0.046	0.046
	0.8	0.295	0.100	0.063	0.054	0.016	0.037	0.045	0.045
	0.9	0.171	0.079	0.052	0.047	0.020	0.043	0.049	0.048

Table 2: Rejection rates of the CUSUM change-point tests for stochastic volatility time series of length n which satisfy (14). The calculations are based on 5,000 simulation runs.

			$h = 0.25$				$h = 0.5$				$h = 1$			
	τ	H	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$
CUSUM	0.25	0.6	1.000	0.773	0.178	0.070	1.000	0.959	0.296	0.075	1.000	0.996	0.530	0.089
		0.7	0.994	0.506	0.118	0.060	1.000	0.769	0.159	0.061	1.000	0.945	0.279	0.074
		0.8	0.953	0.295	0.082	0.057	0.998	0.516	0.105	0.057	1.000	0.752	0.159	0.064
		0.9	0.889	0.200	0.068	0.053	0.986	0.363	0.083	0.055	0.998	0.575	0.113	0.047
	0.5	0.6	1.000	0.912	0.253	0.073	1.000	0.995	0.448	0.092	1.000	1.000	0.739	0.150
		0.7	1.000	0.698	0.140	0.065	1.000	0.926	0.239	0.067	1.000	0.994	0.439	0.092
		0.8	0.994	0.448	0.096	0.060	1.000	0.724	0.149	0.059	1.000	0.914	0.248	0.069
		0.9	0.972	0.328	0.077	0.047	0.999	0.573	0.110	0.050	1.000	0.803	0.181	0.065
self-norm. CUSUM	0.25	0.6	0.773	0.247	0.067	0.054	0.979	0.544	0.131	0.055	0.998	0.818	0.251	0.066
		0.7	0.583	0.140	0.063	0.055	0.883	0.314	0.085	0.051	0.973	0.530	0.141	0.057
		0.8	0.347	0.078	0.046	0.044	0.639	0.151	0.060	0.047	0.828	0.278	0.082	0.041
		0.9	0.252	0.052	0.049	0.052	0.513	0.111	0.059	0.050	0.683	0.189	0.065	0.048
	0.5	0.6	0.896	0.415	0.110	0.050	0.990	0.773	0.216	0.066	0.998	0.948	0.443	0.098
		0.7	0.782	0.247	0.067	0.055	0.956	0.528	0.123	0.054	0.991	0.796	0.243	0.072
		0.8	0.579	0.141	0.046	0.046	0.842	0.294	0.081	0.051	0.946	0.531	0.126	0.049
		0.9	0.450	0.107	0.057	0.044	0.743	0.230	0.068	0.050	0.879	0.394	0.105	0.062

Table 3: Rejection rates of the CUSUM change-point test for stochastic volatility time series of length $n = 300$ with Hurst parameter H and a change-point in the tail parameter α of height h after a proportion τ. The calculations are based on 5,000 simulation runs.

7 Appendix: Proof of Lemma 3.6

Prior to the proof of Lemma 3.6 we establish the following result:

Lemma 7.1. *Let $x < y$ and define*

$$\Delta_n(t; x, y) := M_n(y, t) - M_n(x, t).$$

Then, $\mathrm{E}(M_n(x, t)) = 0$, $\mathrm{E}(\Delta_n(t; x, y)) = 0$, *and the following inequalities hold:*

$$\mathrm{Var}\left(M_n(x, t)\right) \leq \lfloor nt \rfloor F_{\psi(X_1)}(x), \quad \mathrm{Var}\left(\Delta_n(t; x, y)\right) \leq \lfloor nt \rfloor \left(F_{\psi(X_1)}(y) - F_{\psi(X_1)}(x)\right).$$

Proof. We have

$$\Delta_n(t; x, y) = \sum_{j=1}^{\lfloor nt \rfloor} \alpha_j(x, y),$$

where

$$\alpha_j(x, y) := 1_{\{x < \psi(X_j) \leq y\}} - \mathrm{E}\left(1_{\{x < \psi(X_j) \leq y\}} \mid \mathcal{F}_{j-1}\right). \tag{15}$$

Obviously,

$$\mathrm{E}\left(\sum_{j=1}^{\lfloor nt \rfloor} \alpha_j(x, y)\right) = 0.$$

Furthermore,

$$\mathrm{Var}\left(\sum_{j=1}^{\lfloor nt \rfloor} \alpha_j(x, y)\right) = \mathrm{E}\left[\left(\sum_{j=1}^{\lfloor nt \rfloor} \alpha_j(x, y)\right)^2\right] = \sum_{i=1}^{\lfloor nt \rfloor} \sum_{j=1}^{\lfloor nt \rfloor} \mathrm{E}\left(\alpha_i(x, y)\alpha_j(x, y)\right).$$

Note that for $i < j$ we have

$$\begin{aligned}
&\mathrm{E}\left(\alpha_i(x, y)\alpha_j(x, y)\right) \\
&= \mathrm{E}\left[\left(1_{\{x<\psi(X_i)\leq y\}} - \mathrm{E}\left(1_{\{x<\psi(X_i)\leq y\}} \mid \mathcal{F}_{i-1}\right)\right)\left(1_{\{x<\psi(X_j)\leq y\}} - \mathrm{E}\left(1_{\{x<\psi(X_j)\leq y\}} \mid \mathcal{F}_{j-1}\right)\right)\right] \\
&= \mathrm{E}\left(1_{\{x<\psi(X_i)\leq y\}} 1_{\{x<\psi(X_j)\leq y\}}\right) + \mathrm{E}\left(\mathrm{E}\left(1_{\{x<\psi(X_i)\leq y\}} \mid \mathcal{F}_{i-1}\right) \mathrm{E}\left(1_{\{x<\psi(X_j)\leq y\}} \mid \mathcal{F}_{j-1}\right)\right) \\
&\quad - \mathrm{E}\left(1_{\{x<\psi(X_i)\leq y\}} \mathrm{E}\left(1_{\{x<\psi(X_j)\leq y\}} \mid \mathcal{F}_{j-1}\right)\right) - \mathrm{E}\left(1_{\{x<\psi(X_j)\leq y\}} \mathrm{E}\left(1_{\{x<\psi(X_i)\leq y\}} \mid \mathcal{F}_{i-1}\right)\right) \\
&= \mathrm{E}\left(1_{\{x<\psi(X_i)\leq y\}} 1_{\{x<\psi(X_j)\leq y\}}\right) + \mathrm{E}\left(\mathrm{E}\left(1_{\{x<\psi(X_j)\leq y\}} \mathrm{E}\left(1_{\{x<\psi(X_i)\leq y\}} \mid \mathcal{F}_{i-1}\right) \mid \mathcal{F}_{j-1}\right)\right) \\
&\quad - \mathrm{E}\left(\mathrm{E}\left(1_{\{x<\psi(X_i)\leq y\}} 1_{\{x<\psi(X_j)\leq y\}} \mid \mathcal{F}_{j-1}\right)\right) - \mathrm{E}\left(1_{\{x<\psi(X_j)\leq y\}} \mathrm{E}\left(1_{\{x<\psi(X_i)\leq y\}} \mid \mathcal{F}_{i-1}\right)\right) \\
&= \mathrm{E}\left(1_{\{x<\psi(X_i)\leq y\}} 1_{\{x<\psi(X_j)\leq y\}}\right) + \mathrm{E}\left(1_{\{x<\psi(X_j)\leq y\}} \mathrm{E}\left(1_{\{x<\psi(X_i)\leq y\}} \mid \mathcal{F}_{i-1}\right)\right) \\
&\quad - \mathrm{E}\left(1_{\{x<\psi(X_i)\leq y\}} 1_{\{x<\psi(X_j)\leq y\}}\right) - \mathrm{E}\left(1_{\{x<\psi(X_j)\leq y\}} \mathrm{E}\left(1_{\{x<\psi(X_i)\leq y\}} \mid \mathcal{F}_{i-1}\right)\right) \\
&= 0.
\end{aligned}$$

Due to stationarity we have

$$\begin{aligned}\mathrm{E}\left(\alpha_i^2(x,y)\right) &= \mathrm{E}\left[\left(1_{\{x<\psi(X_i)\le y\}} - \mathrm{E}\left(1_{\{x<\psi(X_i)\le y\}} \mid \mathcal{F}_{i-1}\right)\right)^2\right] \\ &= \mathrm{E}\left[1_{\{x<\psi(X_i)\le y\}}\right] - \mathrm{E}\left[1_{\{x<\psi(X_i)\le y\}}\,\mathrm{E}\left(1_{\{x<\psi(X_i)\le y\}} \mid \mathcal{F}_{i-1}\right)\right] \\ &\le F_{\psi(X_1)}(y) - F_{\psi(X_1)}(x).\end{aligned}$$

It follows that

$$\mathrm{Var}\left(\sum_{j=1}^{\lfloor nt\rfloor}\alpha_j(x,y)\right) = \sum_{j=1}^{\lfloor nt\rfloor}\mathrm{E}\left(\alpha_j^2(x,y)\right) \le \lfloor nt\rfloor\left(F_{\psi(X_1)}(y) - F_{\psi(X_1)}(x)\right).$$

□

Proof of Lemma 3.6. In order to prove Lemma 3.6 we have to verify tightness in $\mathbb{D}([-\infty,\infty]\times[0,1])$. For this we quote Theorem 3 in [14], adapted to $\mathbb{D}([-\infty,\infty]\times[0,1])$, i.e. we assume that the following theorem holds:

Theorem 7.2. *Let $X_n(x,t)$ be a sequence of random elements of $\mathbb{D}([-\infty,\infty]\times[0,1])$ and let T_n^k denote a natural k-stopping time for $X_n(x,t)$. If $X_n(x,t)$ is tight on the line for each $(x,t)\in[-\infty,\infty]\times[0,1]$ and if for all $\delta_n\searrow 0$, all $C>0$ and all uniformly bounded T_n^k*

$$\sup_{0\le t\le 1}\left|X_n(T_n^k+\delta_n,t) - X_n(T_n^k,t)\right| \xrightarrow{P} 0, \text{ as } n\to\infty, \tag{16}$$

$$\sup_{-C\le x\le C}\left|X_n(x,T_n^k+\delta_n) - X_n(x,T_n^k)\right| \xrightarrow{P} 0, \text{ as } n\to\infty, \tag{17}$$

then X_n is tight in $\mathbb{D}([-\infty,\infty]\times[0,1])$.

In the following we will show that the conditions of the theorem hold for $X_n(x,t) := n^{-1/2}M_n(x,t)$. Then, Lemma 3.6 immediately follows from an application of Theorem 7.2. Indeed, due to Lemma 3.5 $n^{-1/2}M_n(x,t)$ is tight for each $(x,t)\in(-\infty,\infty)\times[0,1]$. Therefore, it remains to show (16) and (17).

Let T_n^k denote a natural, uniformly bounded k-stopping time for $X_n(x,t)$. Thus, $|T_n^k|\le M$ for some $M>0$ and for all n. Define $\tau_n := \lfloor nT_n^k\rfloor$.

Note that

$$\left|\frac{1}{\sqrt{n}}M_n(x,T_n^k+\delta_n) - \frac{1}{\sqrt{n}}M_n(x,T_n^k)\right| = \left|\frac{1}{\sqrt{n}}\sum_{j=\tau_n+1}^{\tau_n+\lfloor n\delta_n\rfloor}\xi_j(x,\omega)\right|,$$

where $\xi_j(x,\omega) := (\xi_j(x))(\omega)$. For $C > 0$ we have

$$\begin{aligned}
&P\left(\left\{\omega : \sup_{-C\le x\le C}\left|\frac{1}{\sqrt{n}}\sum_{j=\tau_n(\omega)+1}^{\tau_n(\omega)+\lfloor n\delta_n\rfloor}\xi_j(x,\omega)\right| > \varepsilon\right\}\right)\\
&= \sum_{m=0}^{M} P\left(\left\{\omega : \sup_{-C\le x\le C}\left|\frac{1}{\sqrt{n}}\sum_{j=\tau_n(\omega)+1}^{\tau_n(\omega)+\lfloor n\delta_n\rfloor}\xi_j(x,\omega)\right| > \varepsilon\right\} \cap \{\omega \mid \tau_n(\omega) = m\}\right)\\
&\le \sum_{m=0}^{M} P\left(\left\{\omega : \sup_{-C\le x\le C}\left|\frac{1}{\sqrt{n}}\sum_{j=m+1}^{m+\lfloor n\delta_n\rfloor}\xi_j(x,\omega)\right| > \varepsilon\right\}\right)\\
&= \sum_{m=0}^{M} P\left(\left\{\omega : \sup_{-C\le x\le C}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor n\delta_n\rfloor}\xi_j(x,\omega)\right| > \varepsilon\right\}\right)\\
&= (M+1)P\left(\sup_{-C\le x\le C}\left|\frac{1}{\sqrt{n}}M_n(x,\delta_n)\right| > \varepsilon\right).
\end{aligned}$$

Therefore, to show (17), it suffices to prove that

$$\sup_{-C\le x\le C}\left|\tilde{M}_n(x)\right| \xrightarrow{P} 0, \text{ as } n\to\infty, \tag{18}$$

with

$$\tilde{M}_n(x) := \frac{1}{\sqrt{n}}M_n(x,\delta_n).$$

Due to Lemma 7.1, $\mathrm{E}(\tilde{M}_n(x)) = 0$ and $\mathrm{Var}(\tilde{M}_n(x)) \le \delta_n F_{\psi(X_1)}(x) \longrightarrow 0$, hence $\tilde{M}_n(x)$ converges to 0 in probability. This implies fidi-convergence of $\tilde{M}_n$ as a process with values in $\mathbb{D}[-C,C]$. In order to show tightness of $\tilde{M}_n$, we adopt the argument that proves Theorem 15.6 in [5]. For any function v in $\mathbb{D}[-C,C]$ define the modulus $\omega_v''(\delta)$ by

$$\omega_v''(\delta) = \sup\min\{|v(t)-v(t_1)|, |v(t_2)-v(t)|\},$$

where the supremum extends over t_1, t, and t_2 satisfying

$$t_1 \le t \le t_2,\ t_2 - t_1 \le \delta.$$

Under the assumption of fidi-convergence it suffices to show that for each positive ε and η, there exists a δ, $0<\delta<1$, and an integer n_0 such that

$$P_n\left(\omega''_{\tilde{M}_n}(\delta) \ge \varepsilon\right) \le \eta,\ n \ge n_0,$$

(see Theorem 15.4 in [5]).

Define

$$M_m'' := \max_{0 \le i \le j \le k \le m} \min \{|S_j - S_i|, |S_k - S_j|\},$$

where $S_i = \tilde{M}_n(\tau + \frac{i}{m}\delta)$. By Theorem 12.5 in [5]

$$P\left(M_m'' \ge \lambda\right) \le \frac{K}{\lambda^2}(u_1 + \ldots + u_m)^2 \tag{19}$$

holds for all positive λ and some constant K, if

$$P(\{|S_j - S_i| \ge \lambda, |S_k - S_j| \ge \lambda\}) \le \frac{1}{\lambda^2}\left(\sum_{i<l\le k} u_l\right), \; 0 \le i \le j \le k \le m.$$

For $x_1 \le x \le x_2$ we have

$$\left|\tilde{M}_n(x_1) - \tilde{M}_n(x)\right| = \frac{1}{\sqrt{n}}\left|\sum_{j=1}^{\lfloor n\delta_n \rfloor} \alpha_j(x_1, x)\right|,$$

$$\left|\tilde{M}_n(x) - \tilde{M}_n(x_2)\right| = \frac{1}{\sqrt{n}}\left|\sum_{j=1}^{\lfloor n\delta_n \rfloor} \alpha_j(x, x_2)\right|$$

with α_j defined by (15). Because of the Cauchy - Schwarz inequality for expected values and Lemma 7.1 it follows that

$$\begin{aligned}
&\mathrm{E}\left[\frac{1}{\sqrt{n}}\left|\sum_{j=1}^{\lfloor n\delta_n \rfloor} \alpha_j(x_1, x)\right| \frac{1}{\sqrt{n}}\left|\sum_{j=1}^{\lfloor n\delta_n \rfloor} \alpha_j(x, x_2)\right|\right] \\
&\le \sqrt{\frac{1}{n}\mathrm{E}\left[\left(\sum_{j=1}^{\lfloor n\delta_n \rfloor} \alpha_j(x_1, x)\right)^2\right] \frac{1}{n}\mathrm{E}\left[\left(\sum_{j=1}^{\lfloor n\delta_n \rfloor} \alpha_j(x, x_2)\right)^2\right]} \le \delta_n \left(F_{\psi(X_1)}(x_2) - F_{\psi(X_1)}(x_1)\right).
\end{aligned} \tag{20}$$

The Markov inequality yields

$$\begin{aligned}
P\left(|S_j - S_i| \ge \lambda, |S_k - S_j| \ge \lambda\right) &\le P\left(|S_j - S_i|\,|S_k - S_j| \ge \lambda^2\right) \\
&\le \frac{1}{\lambda^2}\mathrm{E}\,|S_j - S_i|\,|S_k - S_j|.
\end{aligned}$$

Therefore, it follows by (20) that

$$P\left(|S_j - S_i| \ge \lambda, |S_k - S_j| \ge \lambda\right) \le \frac{1}{\lambda^2}\sum_{l=i+1}^{k} u_l$$

with

$$u_l := \delta_n \left(F_{\psi(X_1)} \left(\tau + \frac{l}{m}\delta \right) - F_{\psi(X_1)} \left(\tau + \frac{l-1}{m}\delta \right) \right).$$

As a result, we have

$$P\left(M_m^{''} \geq \varepsilon\right) \leq \frac{K}{\varepsilon^2} \sum_{l=1}^{m} u_l = \frac{K}{\varepsilon^2}\delta_n \left(F_{\psi(X_1)}(\tau+\delta) - F_{\psi(X_1)}(\tau) \right).$$

Define

$$\omega^{''}(\tilde{M}_n, [\tau, \tau+\delta]) := \sup \min \left\{ |\tilde{M}_n(t) - \tilde{M}_n(t_1)|, |\tilde{M}_n(t_2) - \tilde{M}_n(t)| \right\},$$

where the supremum extends over t_1, t, t_2 satisfying $\tau \leq t_1 \leq t \leq t_2 \leq \tau + \delta$. Letting $m \longrightarrow \infty$ in (19) yields

$$P\left(\omega^{''}(\tilde{M}_n, [\tau, \tau+\delta]) \geq \varepsilon\right) \leq \frac{K}{\varepsilon^2}\delta_n \left(F_{\psi(X_1)}(\tau+\delta) - F_{\psi(X_1)}(\tau) \right) \tag{21}$$

due to right-continuity of $\tilde{M}_n$. Suppose that $\delta = \frac{C}{u}$ for some integer u and assume that

$$\omega^{''}(\tilde{M}_n, [-C + 2i\delta, -C + (2i+2)\delta]) \leq \varepsilon, \ 0 \leq i \leq u-1, \tag{22}$$

$$\omega^{''}(\tilde{M}_n, [-C + (2i+1)\delta, -C + (2i+3)\delta]) \leq \varepsilon, \ 0 \leq i \leq u-2. \tag{23}$$

If $t_1 \leq t \leq t_2$ and $t_2 - t_1 \leq \delta$, then t_1 and t_2 both lie in one of the $2u-1$ intervals $[-C + 2i\delta, -C + (2i+2)\delta]$, $0 \leq i \leq u-1$, $[-C + (2i+1)\delta, -C + (2i+3)\delta]$, $0 \leq i \leq u-2$, so that

$$\min \left\{ \left|\tilde{M}_n(t) - \tilde{M}_n(t_1)\right|, \left|\tilde{M}_n(t_2) - \tilde{M}_n(t)\right| \right\} \leq \varepsilon.$$

Thus, (22) and (23) together imply $\omega^{''}_{\tilde{M}_n}(\delta) \leq \varepsilon$. It now follows by (21) that

$$P\left(\omega^{''}_{\tilde{M}_n}(\delta) \geq \varepsilon\right) \leq \frac{K}{\varepsilon^2}\delta_n \left(\Sigma' + \Sigma'' \right),$$

where each of Σ' and Σ'' is a sum of the form

$$\sum_{k=1}^{r} \left(F_{\psi(X_1)}(t_k) - F_{\psi(X_1)}(t_{k-1}) \right)$$

with $-C \leq t_1 \leq \ldots \leq t_r \leq C$ and $t_k - t_{k-1} \leq 2\delta$. Hence, we may conclude that

$$P\left(\omega^{''}_{\tilde{M}_n}(\delta) \geq \varepsilon\right) \leq \frac{2K}{\varepsilon^2}\delta_n \left(F_{\psi(X_1)}(1) - F_{\psi(X_1)}(0) \right).$$

Since the right-hand side of the above inequality converges to 0, (17) has been proved.

Thus, it remains to show (16). Note that

$$\sup_{0\le t\le 1}\left|X_n(T_n^k+\delta_n,t)-X_n(T_n^k,t)\right| = \sup_{0\le t\le 1}|M_n^*(t)|$$

with

$$M_n^*(t) := \frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor nt\rfloor}\left(1_{\{0\le\psi(X_j)\le\delta_n\}} - \mathrm{E}\left(1_{\{0\le\psi(X_j)\le\delta_n\}}\mid\mathcal{F}_{j-1}\right)\right).$$

Lemma 7.1 yields $\mathrm{E}(M_n^*(t))=0$ and $\mathrm{Var}(M_n^*(t))\le t\left(F_{\psi(X_1)}(\delta_n)-F_{\psi(X_1)}(0)\right)$. Thus, $M_n^*(t)$ converges to 0 in probability to continuity of $F_{\psi(X_1)}$. This implies fidi-convergence of M_n^* as a process with values in $\mathbb{D}[0,1]$. Again, we make use of Theorem 15.4 in [5], i.e. we verify that for each positive ε and η, there exists a δ, $0<\delta<1$, and an integer n_0 such that

$$P_n\left(\omega''_{M_n^*}(\delta)\ge\varepsilon\right)\le\eta,\ n\ge n_0.$$

Define

$$M''_m := \max_{0\le i\le j\le k\le m}\min\left\{|S_j-S_i|,|S_k-S_j|\right\},$$

where $S_i = M_n^*(\tau+\frac{i}{m}\delta)$.

We have

$$|M_n^*(t_1)-M_n^*(t)| = \frac{1}{\sqrt{n}}\left|\sum_{j=\lfloor nt_1\rfloor+1}^{\lfloor nt\rfloor}\alpha_j(0,\delta_n)\right|,$$

$$|M_n^*(t)-M_n^*(t_2)| = \frac{1}{\sqrt{n}}\left|\sum_{j=\lfloor nt\rfloor+1}^{\lfloor nt_2\rfloor}\alpha_j(0,\delta_n)\right|,$$

where α_j as defined before. The Cauchy-Schwarz inequality yields

$$\begin{aligned}
&\mathrm{E}\,|M_n^*(t_1)-M_n^*(t)|\,|M_n^*(t)-M_n^*(t_2)|\\
&=\mathrm{E}\left[\frac{1}{\sqrt{n}}\left|\sum_{j=\lfloor nt_1\rfloor+1}^{\lfloor nt\rfloor}\alpha_j(x_1,x)\right|\frac{1}{\sqrt{n}}\left|\sum_{j=\lfloor nt\rfloor+1}^{\lfloor nt_2\rfloor}\alpha_j(x,x_2)\right|\right]\\
&\le\sqrt{(t-t_1)(t_2-t)}\left(F_{\psi(X_1)}(\delta_n)-F_{\psi(X_1)}(0)\right)\\
&\le(t_2-t_1)\left(F_{\psi(X_1)}(\delta_n)-F_{\psi(X_1)}(0)\right).
\end{aligned}$$

By the same argument as in the proof of (17) it follows that

$$P\left(|S_j-S_i|\ge\lambda,|S_k-S_j|\ge\lambda\right)\le\frac{1}{\lambda^2}\sum_{l=i+1}^{k}u_l,$$

with $u_l := \left(F_{\psi(X_1)}(\delta_n) - F_{\psi(X_1)}(0)\right) \frac{1}{m}\delta$.

Therefore, Theorem 12.5 in [5] yields

$$P\left(M_m'' \geq \varepsilon\right) \leq \frac{K}{\varepsilon^2}\delta\left(F_{\psi(X_1)}(\delta_n) - F_{\psi(X_1)}(0)\right).$$

Taking the right-continuity of M_n^* into consideration, we may, as before, conclude that

$$P\left(\omega''_{M_n^*}(\delta) \geq \varepsilon\right) \leq \frac{K}{\varepsilon^2}\delta\left(F_{\psi(X_1)}(\delta_n) - F_{\psi(X_1)}(0)\right).$$

Due to continuity of $F_{\psi(X_1)}$ the right-hand side of the above inequality vanishes as n tends to ∞. This concludes the proof of (16) as well as the proof of Lemma 3.6.

□

References

[1] Donald W. K. Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61:821–856, 1993.

[2] Jan Beran, Yuanhua Feng, Sucharita Ghosh, and Rafal Kulik. *Long memory processes.* Springer, 2013.

[3] I. Berkes, L. Horváth, P. Kokoszka, and Q.-M. Shao. On discriminating between long-range dependence and changes in mean. *Annals of Statistics*, 34(3):1140–1165, 2006.

[4] Annika Betken. Testing for change-points in long-range dependent time series by means of a self-normalized wilcoxon test. *Journal of Time Series Analysis*, 37(6):721–864, 2016.

[5] Patrick Billingsley. *Convergence of Probability Measures.* John Wiley & Sons, Inc., 1968.

[6] Patrick Billingsley. *Convergence of Probability Measures.* John Wiley & Sons, Inc., 1999.

[7] F. Jay Breidt, Nuno Crato, and Pedro de Lima. The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, 83(1-2):325–348, 1998.

[8] M. Csörgö and L. Horvath. *Limit Theorems in Change-Point Analysis.* Wiley, New York., 1997.

[9] Herold Dehling, Aeneas Rooch, and Murad S. Taqqu. Non-parametric change-point tests for long-range dependent data. *Scandinavian Journal of Statistics*, 40(1):153—173, 2013.

[10] Herold Dehling and Murad Taqqu. The empirical process of some long-range dependent sequences with an application to U-statistics. *The Annals of Statistics*, 17(4):1767—1783, 1989.

[11] Aryeh Dvoretzky. Asymptotic normality for sums of dependent random variables. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, 2:513–535, 1972.

[12] Liudas Giraitis, Remigijus Leipus, and Donatas Surgailis. The change-point problem for dependent observations. *Journal of Statistical Planning and Inference*, 53:297 – 310, 1996.

[13] L. Horváth and P. Kokoszka. The effect of long-range dependence on change-point estimators. *Journal of Statistical Planning and Inference*, 64(1):57–81, 1997.

[14] Gail Ivanoff. Stopping times and tightness for multiparameter martingales. *Statistics & probability letters*, 28(2):111–114, 1996.

[15] Rafał Kulik and Philippe Soulier. The tail empirical process for long memory stochastic volatility sequences. *Stochastic Processes and their Applications*, 121(1):109 – 134, 2011.

[16] Ignacio N. Lobato. Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association*, 96:1066–1076, 2001.

[17] Sidney I. Resnick. *Heavy-tail phenomena*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2007. Probabilistic and statistical modeling.

[18] Xiaofeng Shao. A simple test of changes in mean in the possible presence of long-range dependence. *Journal of Time Series Analysis*, 32:598 – 606, 2011.

[19] Xiaofeng Shao and Xianyang Zhang. Testing for change points in time series. *Journal of The American Statistical Association*, 105:1228 – 1240, 2010.

[20] Yixiao Sun, Peter C. B. Phillips, and Sainan Jin. Optimal bandwidth selection in heteroskedasticity autocorrelation robust testing. *Econometrica*, 76:175–194, 2008.

[21] Murad S. Taqqu. Convergence of integrated processes of arbitrary Hermite rank. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 50(1):53–83, 1979.

[22] Johannes Tewes. Change-point tests under local alternatives for long-range dependent processes. *arXiv preprint arXiv:1506.07296*, 2015.