Electronic Journal of Statistics
Vol. 11 (2017) 1600–1659
ISSN: 1935-7524
DOI: 10.1214/17-EJS1262

Detecting long-range dependence in non-stationary time series

Holger Dette, Philip Preuss and Kemal Sen

Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
e-mail: holger.dette@ruhr-uni-bochum.de
philip.preuss@ruhr-uni-bochum.de; kemal.sen@ruhr-uni-bochum.de

Abstract: An important problem in time series analysis is the discrimination between non-stationarity and long-range dependence. Most of the literature considers the problem of testing specific parametric hypotheses of non-stationarity (such as a change in the mean) against long-range dependent stationary alternatives. In this paper we suggest a simple approach, which can be used to test the null-hypothesis of a general non-stationary short-memory against the alternative of a non-stationary long-memory process. The test procedure works in the spectral domain and uses a sequence of approximating tvFARIMA models to estimate the time varying long-range dependence parameter. We prove uniform consistency of this estimate and asymptotic normality of an averaged version. These results yield a simple test (based on the quantiles of the standard normal distribution), and it is demonstrated in a simulation study that - despite of its semi-parametric nature - the new test outperforms the currently available methods, which are constructed to discriminate between specific parametric hypotheses of non-stationarity short- and stationarity long-range dependence.

MSC 2010 subject classifications: Primary 62M10, 62M15; secondary 62G10.
Keywords and phrases: Spectral density, long-memory, non-stationary processes, goodness-of-fit tests, empirical spectral measure, integrated periodogram, locally stationary process, approximating models.

Received July 2016.

1. Introduction

Many time series [like asset volatility or regional temperatures] exhibit a slow decay in the sample autocorrelation function and simple stationary short-memory models can not be used to analyze this type of data. A typical example is displayed in Figure 1, which shows 2048 log-returns of the IBM stock between July 15th 2005 and August 30th 2013, with estimated autocovariance function of the squared returns X_t^2. In this example the assumption of stationarity with a summable sequence of autocovariances, say $(\gamma(k))_{k\in\mathbb{N}}$, is hard to justify for the volatility process. Long-range dependent processes have been introduced as an attractive alternative to model features of this type using an autocovariance function with the property

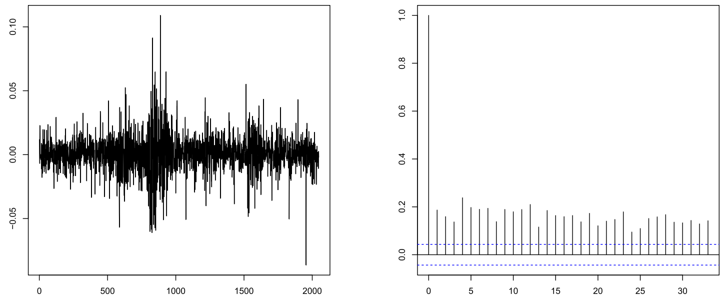

Fig 1. *Left panel: log-returns of the IBM stock between July 15th 2005 and August 30th 2013; right panel: Sample autocovariance function of the squared returns* X_t^2

$$\gamma(k) \sim Ck^{2d-1}$$

as $k \to \infty$, where $d \in (0, 0.5)$ denotes a "long memory" parameter. Statistical models (and corresponding theory) for long-range dependent processes are very well developed [see Doukhan et al. (2003) or Palma (2007) for recent surveys] and have found applications in numerous fields [see Breidt et al. (1998), Beran et al. (2006) or Haslett and Raftery (1989) for such an approach in the framework of asset volatility, video traffic and wind power modeling]. However, it was pointed out by several authors that the observation of "long memory" features in the sample autocovariance function can be as well explained by non stationarity [see Mikosch and Starica (2004) or Chen et al. (2010) among many others]. This is clearly demonstrated in Figure 2, which shows the sample autocovariances of the squared returns from a fit of the (non-stationary) model $X_{t,T} = \sigma(t/T)Z_t$ for the returns [here Z_t is an i.i.d. sequence and $\sigma(\cdot)$ is piecewise-constant, cf. Starica and Granger (2005) or Fryzlewicz et al. (2006) for more details], and from a stationary FARIMA$(3, d, 0)$-fit for the squared ones X_t^2. Both models are able to explain the observed effect of 'long-range dependence' for the volatility process. So, in summary, the same effect can be explained by two completely different modeling approaches. For this reason several authors have pointed out the importance to distinguish between long-memory and non-stationarity [see Starica and Granger (2005), Perron and Qu (2010) or Chen et al. (2010) to mention only a few]. However, there exists a surprisingly small number of statistical procedures which address problems of this type. To the best of our knowledge, Künsch (1986) is the first reference investigating the existence of "long memory" if non-stationarities appear in the time series. In this article a procedure to discriminate between a long-range dependent model and a process with a monotone mean functional and weakly dependent innovations is derived. Later on, Heyde and Dai (1996) developed a method for distinguishing between long-memory and small trends. Sibbertsen and Kruse (2009) tested the null hypothesis of a constant long-memory parameter against a break in the long-memory parameter. Furthermore, Berkes et al. (2006), Baek and Pipiras (2012) and Yau and Davis (2012) investigated CUSUM and likelihood ratio tests to

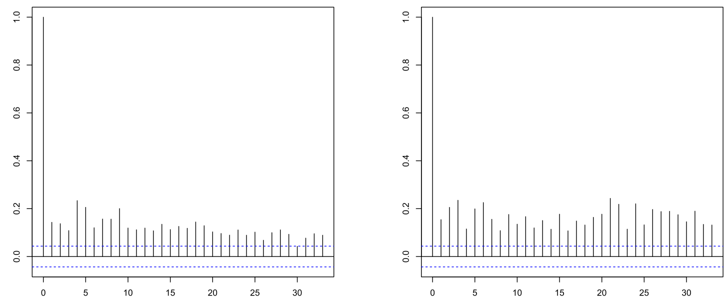

Fig 2. *Left panel: Sample autocovariance function of a simulated time series from a FARIMA(3,d,0)-fit to the* 2048 *squared IBM-returns* X_t^2*, right panel: Sample autocovariance function of* X_t^2 *for* X_t *simulated from the model* $X_{t,T} = \hat{\sigma}(t/T)Z_t$ *with* $\hat{\sigma}(\cdot)$ *estimated by a rolling-window of length* 128.

discriminate between the null hypothesis of no long-range and weak dependence with one change point in the mean.

Although the procedures proposed in these articles are technically mature and work rather well in suitable situations, they are, however, only designed to discriminate between long-range dependence and a very specific change in the first-order structure, like one structural break and two stationary segments of the series. This is rather restrictive, since the expectation might change in a different way than assumed [there could be, for example, continuous changes or multiple breaks instead of a single one] and the second-order structure could be time-varying as well. However, if these or more general non-stationarities occur, the discrimination techniques, which have been proposed in the literature so far, usually fail, and a procedure which is working under less restrictive assumptions is still missing.

The objective of this paper is to fill this gap and to develop a test for the null hypothesis of no long-range dependence in a framework which is flexible enough to deal with different types of non-stationarity in both the first and second-order structure. The general model is introduced in Section 2. Our approach uses an estimate of a (possibly time varying) long-range dependence parameter, which is derived by a sequence of approximating tvFARIMA models with a slightly enlarged parameter space. This statistic estimates a functional which vanishes if and only if the null hypothesis of a short-memory locally stationary process is satisfied. The method is based on some non-intuitive features of averages of unconstrained estimators in models with a constrained parameter space, which become clear from the rather technical proofs given in Section 7. In order to make these phenomena also visible to readers which are less familiar with the technical machinery used for the asymptotic analysis of non-stationary long range dependent processes we provide in Section 3 a motivation of our approach in the context of the classical nonparametric regression model with repeated observations.

In Section 4 we return to the locally stationary long range dependent time series model and prove consistency and asymptotic normality of a corresponding test statistic under the null hypothesis of no long-range dependence. As a consequence we obtain a nonparametric test, which is based on the quantiles of the standard normal distribution and therefore very easy to implement. The finite sample properties of the new test are investigated in Section 5, which also provides a comparison with the competing procedures with a focus on non-stationarities. We demonstrate the superiority of the new method and also illustrate its application in two data examples.

2. Locally stationary long-range dependent processes

In order to develop a test for the presence of long-range dependence which can deal with different kinds of non-stationarity, a set-up is required which includes short-memory processes with a rather general time-varying first and second order structure and a reasonable long-range dependent extension. For this purpose, we consider a triangular scheme $(\{X_{t,T}\}_{t=1,\ldots,T})_{T\in\mathbb{N}}$ of locally stationary long-memory processes, which have an MA(∞) representation of the form

$$X_{t,T} = \mu(t/T) + \sum_{l=0}^{\infty} \psi_{t,T,l} Z_{t-l}, \quad t = 1, \ldots, T, \tag{2.1}$$

where

$$\sup_{T\in\mathbb{N}} \sup_{t\in\{1,\ldots,T\}} \sum_{l=0}^{\infty} \psi_{t,T,l}^2 \quad < \quad \infty, \tag{2.2}$$

$\mu : [0,1] \to \mathbb{R}$ is a trend function and $\{Z_t\}_{t\in\mathbb{Z}}$ are independent standard normal distributed random variables. The assumption of a normal distribution for the innovations is made to simplify the technical arguments in the proofs of our results [see Section 7] and can be replaced by the existence of moments of all order of the random variables Z_t - see Remark 4.8 for more details. Note also that the random variables Z_t have been standardized to have variance 1. Alternatively, one could normalize by $\psi_{t,T,0} = 1$ and allow for an additional parameter in the variance. For the coefficients $\psi_{t,T,l}$ and the function μ in the expansion (2.1) we make the following additional assumptions.

Assumption 2.1. *Let $(\{X_{t,T}\}_{t=1,\ldots,T})_{T\in\mathbb{N}}$ denote a sequence of stochastic processes which have an MA(∞) representation of the form* (2.1) *satisfying* (2.2)*, where μ is twice continuously differentiable. Furthermore, we assume that the following conditions are satisfied:*

1) There exist twice continuously differentiable functions $\psi_l : [0,1] \to \mathbb{R}$ ($l \in \mathbb{Z}$) such that the conditions

$$\sup_{t=1,\ldots,T} \left|\psi_{t,T,l} - \psi_l(t/T)\right| \quad \leq \quad CT^{-1} I(l)^{D-1} \quad \forall l \in \mathbb{N} \tag{2.3}$$

$$\psi_l(u) \quad = \quad a(u) I(l)^{d_0(u)-1} + O(I(l)^{D-2}) \tag{2.4}$$

are satisfied uniformly with respect to $u \in [0,1]$ *as* $l \to \infty$*, where* $I(x) := |x| \cdot 1_{\{x \neq 0\}} + 1_{\{x=0\}}$ *and* $D = \sup_{u\in[0,1]} d_0(u) < 1/2$*. Moreover, the functions* $a : [0,1] \to \mathbb{R}$*,* $d_0 : [0,1] \to [0,1/2)$ *in* (2.4) *are twice continuously differentiable.*

2) *The time varying spectral density* $f : [0,1] \times [-\pi,\pi] \to \mathbb{R}_0^+$

$$f(u,\lambda) := \frac{1}{2\pi}\Big|\sum_{l=0}^{\infty} \psi_l(u)\exp(-i\lambda l)\Big|^2 \tag{2.5}$$

can be represented as

$$f(u,\lambda) = |1 - e^{i\lambda}|^{-2d_0(u)} g(u,\lambda), \tag{2.6}$$

where the function g *defined by*

$$g(u,\lambda) := \frac{1}{2\pi}\big|1 + \sum_{j=1}^{\infty} a_{j,0}(u)\exp(-i\lambda j)\big|^{-2} \tag{2.7}$$

is twice continuously differentiable (note that the identities (2.5) *and* (2.6) *define the coefficients* $a_{j,0}(u)$*).*

3) *There exists a constant* $C \in \mathbb{R}^+$*, which is independent of* u *and* λ*, such that for* $l \neq 0$ *the conditions*

$$\begin{aligned}
\sup_{u\in(0,1)} |\psi_l'(u)| &\leq C\log|l||l|^{D-1}, \qquad (2.8)\\
\sup_{u\in(0,1)} |\psi_l''(u)| &\leq C\log^2|l||l|^{D-1},\\
\sup_{u\in(0,1)} \Big|\frac{\partial}{\partial u} f(u,\lambda)\Big| &\leq C|\log(\lambda)||\lambda|^{-2D},\\
\sup_{u\in(0,1)} \Big|\frac{\partial^2}{\partial u^2} f(u,\lambda)\Big| &\leq C\log^2(\lambda)|\lambda|^{-2D}
\end{aligned}$$

are satisfied for all $\lambda \in [-\pi,\pi]$*.*

Similar locally stationary long-range dependent models have been investigated by Beran (2009), Palma and Olea (2010) and Roueff and von Sachs (2011) and Wu and Zhou (2014). It is also worthwhile to mention that in general (2.4) does not imply (2.6) and (2.7) and vice versa conditions (2.6) and (2.7) do not imply (2.4). Therefore, none of the conditions (2.4), (2.6) or (2.7) can be omitted in Assumption 2.1. It is also worthwhile to mention that equations (2.5), (2.6) and (2.7) do not necessarily imply that $\psi_0(u) = 1$ (which follows by a careful inspection of these relations).

Note also hat in contrast to the standard framework of local stationarity introduced by Dahlhaus (1997) and extended to the long-memory case in Palma and Olea (2010), condition (2.3) is much weaker. For example, in contrast to these references the assumptions made here include tvFARIMA(p,d,q)-models as well [see Theorem 2.2 in Preuß and Vetter (2013)]. Moreover, we mention again that the assumption of Gaussianity is only imposed to simplify the technical arguments in the proofs of our main results - see Remark 4.8 for more details. The very specific form of the function g in (2.7) implies that the process $\{X_{t,T}\}_{t=1,\dots,T}$ can be locally approximated by a FARIMA$(\infty, d, 0)$ process in the sense of (2.3). More precisely, we obtain with

$$b_k(u) = \binom{k + d(u) - 1}{k} \quad \text{and} \quad \Big(\sum_{k=0}^{\infty} a_{k,0}(u) z^k\Big)^{-1} = \sum_{k=0}^{\infty} a_{k,0}^{(-1)}(u) z^k \tag{2.9}$$

($a_{0,0} = 1$) the relation

$$\psi_l(u) = \sum_{k=0}^{l} a_{k,0}^{(-1)}(u) b_{l-k}(u)$$

between the approximating functions $\psi_l(u)$ and the time-varying AR-parameters [see the proof of Lemma 3.2 in Kokoszka and Taqqu (1995) for more details]. The relation (2.9) can be used to calculate the coefficients $a_{k,0}^{-1}(u)$ from the functions $a_{k,0}(u)$, i.e.

$$a_{0,0}^{(-1)}(u) = \frac{1}{a_{0,0}(u)}, \; a_{1,0}^{(-1)}(u) = -\frac{a_{1,0}(u)}{a_{0,0}^2(u)}, \quad \dots$$

In order to further visualize some properties of these kinds of locally stationary long-memory models we introduce for every fixed $u \in [0,1]$ the stationary process

$$X_t(u) := \mu(u) + \sum_{l=0}^{\infty} \psi_l(u) Z_{t-l}.$$

One can show that condition (2.4) implies the existence of bounded functions $y_i : [0,1] \to \mathbb{R}^+$ $(i = 1,2)$ such that the approximations

$$|\mathrm{Cov}(X_t(u), X_{t+k}(u))| \sim y_1(u) k^{2d_0(u)-1} \quad \text{as } k \to \infty \tag{2.10}$$

and

$$f(u,\lambda) \sim y_2(u) |\lambda|^{-2d_0(u)} \quad \text{as } \lambda \to 0 \tag{2.11}$$

hold [see Palma and Olea (2010) for details]. Consequently, the autocovariances $\gamma_k(u,k) = \mathrm{Cov}(X_0(u), X_k(u))$ are not absolutely summable if the function $a(u)$ in (2.4) is not vanishing, and in this case the time varying spectral density $f(u,\lambda)$ has a pole at $\lambda = 0$ for any $u \in [0,1]$ for which $d_0(u)$ is positive. Note that in general the statements (2.10) and (2.11) are not equivalent [see Yong (1974) for a discussion of this problem in the stationary case].

In the framework of these long-range dependent locally stationary processes we now investigate the null hypothesis that the time-varying "long memory" parameter $d_0(u)$ vanishes for all $u \in [0,1]$, i.e. there is no long-range dependence in the locally stationary process $X_{t,T}$. The alternative is defined by the property that the function d_0 is nonnegative on the interval $[0,1]$ and positive on a subset of positive Lebesgue measure. Formally, the hypotheses can be formulated as

$$\begin{aligned} & \mathrm{H}_0 : d_0(u) = 0 \quad \forall u \in [0,1] \qquad (2.12) \\ \text{versus} \qquad & \mathrm{H}_1 : d_0(u) \geq 0 \ \forall u \in [0,1] \text{ and } \lambda(\{u \in [0,1] \mid d_0(u) > 0\}) > 0 \ , \end{aligned}$$

where λ denotes the Lebesgue measure. This hypothesis is obviously equivalent to

$$\mathrm{H}_0 : F = 0 \quad \text{versus} \quad \mathrm{H}_1 : F > 0, \qquad (2.13)$$

where the quantity F is defined by the integral

$$F := \int_0^1 d_0(u) du. \qquad (2.14)$$

In Section 4 we will develop a nonparametric estimator of the function d_0 and the integral F. Roughly speaking, the sample size T is decomposed into M blocks with length N (i.e. $T = NM$), where M is some positive integer. We define the corresponding midpoints in both the time and rescaled time domain by $t_j = N(j-1) + N/2$, $u_j = t_j/T$, respectively, and calculate an estimator $\hat{d}_N(u_j)$ of the long range dependence parameter at the point u_j on each of the M blocks (for the exact definition of the estimator see Section 4). The test statistic is then obtained as

$$\hat{F}_T = \frac{1}{M} \sum_{j=1}^{M} \hat{d}_N(u_j) \qquad (2.15)$$

and could be considered as a Riemann sum of the integral $\int_0 \hat{d}_N(u)du$, which approximates the integral in (2.14). We also note that the proofs of our main results in Section 4 require the smoothness of the time varying long range dependence parameter as specified in Assumption 4.1.

Remark 2.2. (*some boundary issues*) Note that for each $u \in [0,1]$ the local long range dependence parameter $d_0(u)$ is a boundary point of the parameter space $[0, 1/2)$ defined by the two hypotheses in (2.12). However, we will not use this property for the construction of the estimates $\hat{d}_N(u_j)$ of the quantities $d_0(u_j)$, which are aggregated in the statistic (2.15). For this purpose we consider a sequence of approximating tvFARIMA$(k, d, 0)$ models, where the parameter $k = k(T)$ converges to infinity as the sample size increases and the corresponding long range dependence parameters are allowed to vary in intervals of the form $[-\gamma_k, 1/2 - \delta_k]$, where $(\gamma_k)_{k\in\mathbb{N}}$. and $(\delta_k)_{k\in\mathbb{N}}$ are positive sequences converging to 0. We will prove in Theorem 4.3 below that this provides a uniformly consistent estimate of the function d_0 and that an average of these statistics provides a consistent and asymptotically normal distributed estimate of the integral F (see Theorem 4.5 and Theorem 4.6 below). As a consequence we obtain a consistent level-α test for the presence of long-range dependence in non-stationary time series by rejecting the null hypothesis for large values of the estimator of $\hat{F}_T$.

On a first glance these properties are surprising because we use unconstrained (i.e. potentially negative) estimators of the long range dependence parameters in the approximating tvFARIMA models to estimate the non-negative function d_0, but the statements become clear from the rather technical arguments given in the proofs of Section 7. The situation is similar to the problem of testing the hypothesis $H_0 : \mu = 0$ versus $H_1 : \mu > 0$ for the mean of a sample of i.i.d. random variables $X_1, \ldots, X_n$. A test which rejects H_0, whenever $\sqrt{n}\overline{X}_n > \hat{\sigma}_n u_{1-\alpha}$ (here $\hat{\sigma}_n$ is an estimator of the variance and $u_{1-\alpha}$ the $(1-\alpha)$-quantile of the standard normal distribution) has asymptotic level α and is consistent. Moreover, in Section 3 we consider an example of testing for a positive signal in a nonparametric regression model and demonstrate that the aggregation of local statistics of the type $\overline{X}_n$ might have substantial advantages compared to the aggregation of local estimators of the from $\max\{\overline{X}_n, 0\}$, which reflect the constraint $\mu \geq 0$ in its definition.

3. Testing for a positive nonparametric signal

In this section we provide some heuristic explanation for the phenomenon described in the previous paragraph, which is also available to readers which are less familiar with the technical machinery used for the asymptotic analysis of non-stationary long range dependent processes. We will also demonstrate that there are situations where more powerful tests can be obtained by ignoring particular constraints in the estimation procedure. This situation occurs in particular if different estimators are aggregated as described in (2.15).

For this purpose we consider the problem of testing the hypothesis of a vanishing regression function against the alternative that the regression function is positive on the interval $[0, 1]$ in the common nonparametric regression model

$$Y_{ji} = \mu(t_j) + \varepsilon_{ji}; \qquad j = 1, \ldots, M; \qquad i = 1, \ldots, N.$$

Here $\varepsilon_{11}, \ldots, \varepsilon_{MN}$ are i.i.d. standard normal distributed (centered) random variables (this assumption is in fact not necessary but makes some of the following arguments much simpler), $t_j = t_{j,M} = j/M$ and μ is a smooth non-negative Lipschitz continuous function on the interval $[0, 1]$. We are interested in testing the hypothesis

$$H_0 : \mu(t) \equiv 0 \qquad \text{versus} \qquad H_1 : \mu(t) > 0 \qquad \text{for all } t \in [0, 1] \,. \tag{3.1}$$

Note that the alternative could also be considered on the set of all non-negative functions which are positive on a subset of positive Lebesgue measure, say $\mathcal{U} \subset [0, 1]$. As this generalization does not change any of the subsequent arguments (only integrals of the form $\int_0^1 \mu(t)dt$ and sums of the form $\frac{1}{M}\sum_{j=1}^{M} \mu(t_j)$ have to be replaced by $\int_{\mathcal{U}} \mu(t)dt$ and $\frac{1}{M\lambda(\mathcal{U})}\sum_{j=1}^{M} 1_{\mathcal{U}}(\frac{j}{M})\mu(t_j))$ (here $\lambda(\mathcal{U})$ denotes the Lebesgue measure of the set $\mathcal{U}$) we restrict ourselves to the case $\mathcal{U} = [0,1]$ for the sake of transparency.

3.1. Tests based on unconstrained estimators

The idea used in Section 4 below for testing hypotheses of this type translates in the nonparametric regression model to the following procedure. We first define (unconstrained) estimators for the values $\mu(t_j)$, that is $\hat{\mu}_j = \frac{1}{N}\sum_{i=1}^{N} Y_{ji}$, ($j = 1, \dots, M$), and then consider the average

$$T_M = \frac{1}{M}\sum_{j=1}^{M} \hat{\mu}_j \ = \ \frac{1}{MN}\sum_{j=1}^{M}\sum_{i=1}^{N} Y_{ji}.$$

Note that $\mathcal{S}_M = \sqrt{N}(T_M - \frac{1}{M}\sum_{j=1}^{M}\mu(t_j)) = \frac{\sqrt{N}}{M}\sum_{j=1}^{M}(\hat{\mu}_j - \mu(t_j))$ is a sum of independent identically distributed random variables with variance $\mathrm{Var}(\mathcal{S}_M) = 1/M$. Consequently, using a central limit theorem for triangular arrays, shows that $\sqrt{M}\,\mathcal{S}_M \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$ as $M \to \infty, N \to \infty$. Moreover, since μ is Lipschitz continuous, this implies

$$\sqrt{MN}\Big(T_M - \int_0^1 \mu(t)dt\Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

whenever $N = o(M)$. Thus a consistent and asymptotic level α test for the hypothesis (3.1) is obtained by rejecting the null hypothesis H_0, whenever

$$\sqrt{MN}T_M > u_{1-\alpha}, \tag{3.2}$$

where $u_{1-\alpha}$ is the $(1-\alpha)$-quantile of the standard normal distribution.

3.2. Tests based on constrained estimators

Alternatively, and - on a first glance - more reasonable strategy is to use a constrained estimator which addresses the boundary condition $\mu(t) \geq 0$. This gives

$$\tilde{\mu}_j = \max\{0, \hat{\mu}_j\}$$

as estimator for $\mu(t_j)$, and we obtain with the notation $\tilde{T}_M = \frac{1}{M}\sum_{j=1}^{M}\tilde{\mu}_j$ the representation

$$\tilde{\mathcal{S}}_M \quad = \quad \sqrt{N}\,\Big(\tilde{T}_M - \frac{1}{M}\sum_{j=1}^{M}\mu(t_j)\Big) = \frac{\sqrt{N}}{M}\Big(\sum_{i=j}^{M} Z_j + \sum_{j=1}^{M}\delta_j\Big), \tag{3.3}$$

where $Z_j = \max(0,\hat{\mu}_j) - \mathbb{E}[\max(0,\hat{\mu}_j)]$, $\delta_j = \mathbb{E}[\max(0,\hat{\mu}_j)] - \mu(t_j)$. Note that $\hat{\mu}_j \sim \mathcal{N}(\mu(t_j), 1/N)$, which yields

$$\delta_j \quad = \quad \frac{1}{\sqrt{2\pi N}}\exp\Big(-\frac{N\mu^2(t_j)}{2}\Big) - \frac{\mu(t_j)}{\sqrt{\pi}}\int_{\mu(t_j)\sqrt{N/2}}^{\infty}\exp(-z^2)dz. \tag{3.4}$$

This term is of order $o(1)$ (exponentially in N and independent of M, provided that $\mu(t) \geq c > 0$ on $[0,1]$). Note also that

$$\begin{aligned}\mathbb{E}[(\max(0,\hat{\mu}_j))^2] &= \mu^2(t_j) + \frac{1}{N} + \frac{\mu(t_j)}{\sqrt{2\pi N}} \exp\Big(-\frac{N\mu^2(t_j)}{2}\Big) \\ &- \frac{1+N\mu^2(t_j)}{N\sqrt{\pi}} \int_{\mu(t_j)\sqrt{N/2}}^{\infty} \exp(-z^2)dz \ .\end{aligned}$$

This gives for the variance of the random variable Z_j

$$\mathbb{E}[Z_j^2] = \mathrm{Var}(\max(0,\hat{\mu}_j)) = \begin{cases} \frac{1}{N}\big(\frac{1}{2} - \frac{1}{2\pi}\big) & \text{if } \mu(t_j) = 0 \\ \frac{1}{N}(1+o(1)) & \text{if } \mu(t_j) > 0 \end{cases}$$

Ljapunoff's central limit theorem now shows that

$$\sqrt{MN/\sigma_N^2}(\tilde{T}_M - \frac{1}{M}\sum_{j=1}^{M} \mu(t_j) - B_{M,N}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

where $\sigma_N^2 = N\mathbb{E}[Z_i^2]$ and

$$B_{M,N} = \frac{1}{M}\sum_{i=j}^{M} \delta_j = \begin{cases} \frac{1}{M}\sum_{j=1}^{M} \delta_j & \text{if } \mu(t) > 0 \text{ for all } t \\ \frac{1}{\sqrt{2\pi N}} & \text{if } \mu(t) = 0 \text{ for all } t \end{cases} .$$

This implies (observing the Lipschitz continuity of the regression function and $N = o(M)$)

$$\sqrt{NM/\sigma_N^2}\Big(\tilde{T}_M - \int_0^1 \mu(t)dt - B_{M,N}\Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Note that the statistic is asymptotically normal distributed, although we average constrained estimators. Under the null hypothesis things are simplifying. In particular we obtain $\sigma_{H_0}^2 = \sigma_{N,H_0}^2 = N\mathbb{E}_{H_0}[Z_i^2] = \frac{1}{2} - \frac{1}{2\pi}$ and a test based on $\tilde{T}_M$ rejects the null hypothesis H_0, whenever

$$\tilde{T}_M > \frac{1}{\sqrt{2\pi N}} + \sigma_{H_0}\frac{u_{1-\alpha}}{\sqrt{MN}} = \frac{1}{\sqrt{2\pi N}} + \sqrt{\frac{1}{2} - \frac{1}{2\pi}}\frac{u_{1-\alpha}}{\sqrt{MN}}. \tag{3.5}$$

This test has asymptotic level α and is consistent. We conclude this section mentioning once again that the assumption of i.i.d. standard normal distributed errors was made to minimize the technical arguments. All statements remain true for arbitrary centered errors which have moments of order 4. This observation is a simple consequence of the central limit theorem, and in the following finite sample comparison we actually use non-normal error distributions.

3.3. A comparison of the two tests

The use of different estimators for the quantities $\mu(t_i)$ yields to the different tests (3.2) and (3.5) the hypotheses in (3.1). Both test statistics have an asymptotic normal distribution under the null hypothesis and the alternative. A finite sample comparison is given in Table 1 where we report simulation results for the functions

$$\mu_1(t) \equiv 0, \tag{3.6}$$

$$\mu_2(t) = 0.1, \tag{3.7}$$

$$\mu_3(t) = 0.1 + 0.1t. \tag{3.8}$$

The sample sizes are $M = N = 20$ and $M = N = 50$ and we use 10000 simulation runs to estimate the rejection probabilities of the tests (3.2) and (3.5). For the distribution of the errors distribution we use a $(\mathcal{X}_5^2 - 5)/\sqrt{10}$ distribution, in order to demonstrate that the previous findings do not depend on the assumption of normal distributed errors. We observe that the test (3.2) based on the statistic T_M (which uses the unconstrained estimators of the regression function) outperforms the method (3.5) which uses the constrained estimators.

A heuristic explanation for the superiority of the unconstrained estimate is as follows: a constrained estimate produces a systematic bias which has to be reflected in the decision rule, such that the test keeps its required nominal level. The bias occurs at every design point and by taking an average, with respect to the design points it accumulates and has an influence on the power of the test.

	$M = N = 20$					
model	(3.6)		(3.7)		(3.8)	
level	5%	10%	5%	10%	5%	10%
test (3.2)	0.052	0.103	0.634	0.764	0.918	0.966
test (3.5)	0.070	0.118	0.576	0.684	0.873	0.926
	$M = N = 50$					
model	(3.6)		(3.7)		(3.8)	
level	5%	10%	5%	10%	5%	10%
test (3.2)	0.056	0.109	0.999	1.000	1.000	1.000
test (3.5)	0.065	0.112	0.997	0.998	1.000	1.000

TABLE 1

Simulated rejection probabilities of the tests (3.2) *and* (3.5) *for the hypothesis* (3.1) *in model* (3.6) - (3.8).

We can also give a more "theoretical" argument for the advantages of the unconstrained approach. Note that for a positive function μ the power of test (3.2) is approximately given by

$$\mathbb{P}_{H_1}\Big(T_M > \frac{u_{1-\alpha}}{\sqrt{MN}}\Big) \approx \Phi\Big(\sqrt{MN}\int_0^1 \mu(t)dt - u_{1-\alpha}\Big)\ ,$$

where Φ denotes the distribution function of the standard normal distribution. This formula is remarkably precise. For example, if $N = M = 20,\ \mu(t) = 0.1$ we obtain for the power of the test (3.2) 0.638, while the result of the simulation is 0.643. Similarly, the power of the test (3.5) is approximately given by

$$\mathbb{P}_{H_1}\Big(\tilde{T}_M > \frac{1}{\sqrt{2\pi N}} + \sigma_{H_0}\frac{u_{1-\alpha}}{\sqrt{MN}}\Big) \approx \Phi\Big(\frac{\sqrt{MN}}{\sigma_N}\int_0^1 \mu(t)dt + r_{N,M}\Big), \qquad (3.9)$$

where the term $r_{N,M}$ is defined by

$$r_{N,M} = \frac{\sqrt{NM}}{\sigma_N}B_{M,N} - \sqrt{\frac{M}{2\pi\sigma_N^2}} - \frac{\sigma_{H_0}}{\sigma_N}u_{1-\alpha}.$$

Now, note that $\sigma_N^2 = 1 + o(1)$ and that $B_{M,N} = o(1)$ of exponential order (uniformly) if $\mu(t) \geq c > 0$ for all $t \in [0,1]$ as $M, N \to \infty$ Consequently, the term $r_{N,M}$ will be negative for reasonable large M, N. It actually diverges to $-\infty$, but at a lower rate as the dominating term $\frac{\sqrt{MN}}{\sigma}\int_0^1 \mu(t)dt$ in (3.9), which converges to ∞. This means that the test (3.2) based on unconstrained estimation is more powerful than the test (3.5), which uses constrained estimation.

A similar argument for the superiority of the test (3.2) based on the unconstrained estimators of the regression function can be given for local alternatives of the form $\mu_{M,N}(t) = c(t)/\sqrt{MN}$, where $c : [0,1] \to \mathbb{R}$ is a Lipschitz continuous function. More precisely, the asymptotic power of the tests (3.2) and (3.5) is given by

$$\Phi\Big(\int_0^1 c(t)dt - u_{1-\alpha}\Big)$$

and

$$\Phi\Big(\int_0^1 c(t)dt/\sqrt{2-2/\pi} - u_{1-\alpha}\Big),$$

respectively. As $\sqrt{2-2\pi} \approx 1.1676 > 1$, it follows that the unconstrained test (3.2) also outperforms the test (3.5) under local alternatives. Exemplarily, we display in Table 2 the power of the two tests under the local alternatives $\mu_{M,N}(t) = (1+t)/\sqrt{MN}$.

	$M = N = 20$		$M = N = 50$		$M = N = 100$		$M = N = 200$		$M = N = 500$	
level	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
(3.2)	0.458	0.600	0.444	0.591	0.442	0.588	0.437	0.587	0.451	0.588
(3.5)	0.420	0.536	0.393	0.516	0.382	0.508	0.378	0.506	0.379	0.509

TABLE 2

Simulated rejection probabilities of the tests (3.2) and (3.5) under local alternatives in model (3.8).

In the following section we will use a similar approach based on averages of unconstrained estimates of the function $d_0(\cdot)$ in sequence of approximating tvFARIMA models. The proofs in Section 7 show that this approach provides a consistent and asymptotic level α test for the hypotheses (2.13).

3.4. Two sided hypotheses

The problem of testing two-sided hypothesis of the form

$$H_0 : \mu(t) \equiv 0 \text{ versus } H_1 : \mu(t) \neq 0 \text{ for all } t \in [0,1] \tag{3.10}$$

can be addressed by rejecting the null hypothesis for large values of the statistic $\hat{A}_M = \frac{1}{M}\sum_{j=1}^{M} |\hat{\mu}_j|$. It can be shown by similar but slightly more complicated arguments as given in the previous paragraphs that

$$\sqrt{MN/\tau_N^2}\Big(\hat{A}_M - \int_0^1 |\mu(t)|dt - C_{M,N}\Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

where $\tau_N^2 = N \text{ Var } (|\hat{\mu}_\mathrm{j}|)$,

$$C_{M,N} = \begin{cases} \frac{2}{M}\sum_{j=1}^{M} \delta_j & \text{if } \mu(t) \neq 0 \text{ for all } t \\ \sqrt{\frac{2}{\pi N}} & \text{if } \mu(t) = 0 \text{ for all } t \end{cases},$$

and the quantities δ_j are defined in (3.4). Thus, the null hypothesis in (3.10) is rejected if

$$\hat{A}_M > \sqrt{\frac{2}{\pi N}} + \tau_{H_0}\frac{u_{1-\alpha}}{\sqrt{MN}}$$

where $\tau_{H_0}^2 = N \text{ Var}_{\mathrm{H_0}}(|\hat{\mu}_\mathrm{j}|) = 1 - 2/\pi$.

4. Testing short- versus long-memory

In order to estimate the integral F we use a sequence of semi-parametric models approximating the processes $\{X_t(u)\}_{t\in\mathbb{Z}}$ with time varying spectral density (2.6) and proceed in several steps. First we choose an increasing sequence $k = k(T) \in \mathbb{N}$, which diverges 'slowly' to infinity as the sample size T grows, and fit a tvFARIMA(k,d,0) model to the data. To be precise, we consider a locally stationary long-memory model with time varying spectral density $f : [0,1] \times [-\pi,\pi] \to \mathbb{R}_0^+$ defined by

$$f_{\theta_k(u)}(\lambda) = |1 - \exp(i\lambda)|^{-2d(u)} g_k(u,\lambda), \tag{4.1}$$

where

$$g_k(u,\lambda) = \frac{1}{2\pi}|1 + \sum_{j=1}^{k} a_j(u)\exp(-i\lambda j)|^{-2}$$

and $\theta_k = (d, a_1, \ldots, a_k) : [0,1] \to \mathbb{R}^{k+1}$ is a vector valued function. We emphasize again that $k = k(T)$ depends on the sample size and refer to Assumption 4.1 for the precise condtions regarding its growth rate. We then estimate the function $\theta_k(u)$ by a localized Whittle-estimator, that is

$$\hat{\theta}_{N,k}(u) \quad = \quad \arg\min_{\theta_k \in \Theta_{\lfloor uT\rfloor/T,k}} \mathcal{L}_{N,k}^{\hat{\mu}}(\theta_k, u), \tag{4.2}$$

where

$$\mathcal{L}_{N,k}^{\hat{\mu}}(\theta_k, u) := \frac{1}{4\pi} \int_{-\pi}^{\pi} \Big(\log(f_{\theta_k}(\lambda)) + \frac{I_N^{\hat{\mu}}(u, \lambda)}{f_{\theta_k}(\lambda)} \Big) d\lambda \tag{4.3}$$

denotes the (local) Whittle likelihood [see Dahlhaus and Giraitis (1998) or Dahlhaus and Polonik (2009)] and for each $u \in [0, 1]$ the parameter space $\Theta_{u,k} \subset \mathbb{R}^{k+1}$ is a compact set which will be specified in Assumption 4.1. In (4.2) and (4.3) the quantity

$$I_N^{\hat{\mu}}(u, \lambda) = \Big| \frac{1}{\sqrt{2\pi N}} \sum_{p=0}^{N-1} \Big[X_{\lfloor uT \rfloor - N/2+1+p,T} - \hat{\mu}(\lfloor uT \rfloor - N/2 + 1 + p, T) \Big] e^{-ip\lambda} \Big|^2, \tag{4.4}$$

denotes the mean-corrected local periodogram, N is an even window-length which is 'small' compared to T and $\hat{\mu}$ is an asymptotically unbiased estimator of the mean function $\mu : [0, 1] \to \mathbb{R}$, see Dahlhaus (1997). Here and throughout this paper we use the convention $X_{j,T} = 0$ for $j \notin \{1, ..., T\}$. We finally obtain an estimator $\hat{d}_N(u)$ for the time-varying long-memory parameter by taking the first component of the $(k + 1)$ dimensional vector $\hat{\theta}_{N,k}(u)$ defined in (4.2). We emphasize that the tvFARIMA models are only used to define the estimator $\hat{d}_N(u)$ as the solution of the optimization problem (4.2).

It will be demonstrated in Theorem 4.3 below that - provided that the "true" underlying process can be approximated by tvFARIMA models - this approach results in a uniformly consistent estimator of the time-varying long-memory parameter. For this purpose we define $\theta_{0,k}(u) := (d_0(u), a_{1,0}(u), ..., a_{k,0}(u))$ as the $(k+1)$ dimensional vector containing the long memory parameter $d_0(u)$ and the first k AR-parameter functions $a_{1,0}(u), ..., a_{k,0}(u)$ of the approximating process $\{X_t(u)\}_{t\in\mathbb{Z}}$ defined by the representation (2.6) and (2.7). Here and throughout this paper, A_{11} denotes the element in the position (1,1) and $\|A\|_{sp}$ the spectral norm of the matrix $A = (a_{ij})_{i,j=1}^k$, respectively. We state the following technical assumptions.

Assumption 4.1. Let $k = k(T)$ be a sequence converging to infinity for increasing sample size T and let $(\gamma_\ell)_{\ell\in\mathbb{N}}$ and $(\delta_\ell)_{\ell\in\mathbb{N}}$ denote positive sequences in the interval $(0, \min\{1/4, 1/2 - D\})$ such that

$$\liminf_{T\to\infty} \gamma_{k(T)} \log T > 0 \ , \quad \liminf_{k\to\infty} \delta_{k(T)} \log T > 0,$$
$$\lim_{T\to\infty} \gamma_{k(T)} = 0 \ , \quad \lim_{k\to\infty} \delta_{k(T)} = 0.$$

For each $u \in [0, 1]$ and $k \in \{k(T), T \in \mathbb{N}\}$ define $\Theta_{u,k} = [-\gamma_k, 1/2-\delta_k] \times \Theta_{u,k,1} \times \ldots \times \Theta_{u,k,k}$, where the constant D is the same as in Assumption 2.1. For each $i = 1, \ldots, k$ $\Theta_{u,k,i}$ is a compact set with a finite number (independent of u, k, i) of connected components with positive Lebesgue measure. Let Θ_k denote the space of all four times continuously differentiable functions $\theta_k : [0, 1] \to \mathbb{R}^{k+1}$ with $\theta_k(u) \in \Theta_{u,k}$ for all $u \in [0, 1]$. If $\theta_k(u)$ and $\theta_k'(u)$ are distinct elements of $\Theta_{u,k}$, we assume that the set $\{\lambda : f_{\theta_k(u)}(\lambda) \neq f_{\theta'_k(u)}(\lambda)\}$ has positive Lebesgue measure. We assume that the following conditions hold for each $k \in \{k(T), T \in \mathbb{N}\}$:

(i) The functions g_k in (4.1) are bounded from below by a positive constant (which is independent of k) and are four times continuously differentiable with respect to λ and u, where all partial derivates of g_k up to the order four are bounded with a constant independent of k.

(ii) For each $u \in [0,1]$ the parameter $\tilde{\theta}_{0,k}(u) = \arg\min_{\theta_k \in \Theta_{u,k}} \mathcal{L}_k(\theta_k, u)$ exists and is uniquely determined, where

$$\mathcal{L}_k(\theta_k, u) := \frac{1}{4\pi} \int_{-\pi}^{\pi} \Big(\log(f_{\theta_k}(\lambda)) + \frac{f(u,\lambda)}{f_{\theta_k}(\lambda)} \Big) d\lambda.$$

Moreover, for each $u \in [0,1]$ the vectors $\tilde{\theta}_{0,k}(u)$ and $\theta_{0,k}(u)$ are interior points of $\Theta_{u,k}$.

(iii) Define

$$\begin{aligned} \Gamma_k(\theta_k) &= \frac{1}{4\pi} \int_{-\pi}^{\pi} f^2_{\theta_k}(\lambda) \nabla f^{-1}_{\theta_k}(\lambda) \nabla f^{-1}_{\theta_k}(\lambda)^T \, d\lambda, \qquad (4.5) \\ V_k(\theta_k, u) &= \frac{1}{4\pi} \int_{-\pi}^{\pi} f^2(u,\lambda) \nabla f^{-1}_{\theta_k}(\lambda) \nabla f^{-1}_{\theta_k}(\lambda)^T d\lambda, \end{aligned}$$

[here ∇ denotes the derivative with respect to the parameter-vector θ_k], then the matrix $\Gamma_k(\theta_{0,k})$ is non-singular for every $u \in [0,1]$, $k \in \{k(T), T \in \mathbb{N}\}$, and

$$\lim_{T\to\infty} \frac{\int_0^1 [\Gamma_k^{-1}(\theta_{0,k}(u))]_{1,1} du}{\int_0^1 [\Gamma_k^{-1}(\theta_{0,k}(u)) V_k(\theta_{0,k}(u), u) \Gamma_k^{-1}(\theta_{0,k}(u))]_{1,1} du} = 1 \qquad (4.6)$$

as $T \to \infty$. Furthermore, condition (4.6) is also satisfied if the function $\theta_{0,k}(u)$ is replaced by any sequence $\tilde{\theta}_T(u)$ such that $\sup_{u\in[0,1]} |\tilde{\theta}_T(u) - \theta_{0,k}(u)| \to 0$. For such a sequence we additionally assume that the condition

$$\lim_{T\to\infty} \frac{\int_0^1 [\Gamma_k^{-1}(\theta_{0,k}(u))]_{1,1} du}{\int_0^1 [\Gamma_k^{-1}(\tilde{\theta}_T(u))]_{1,1} du} = 1$$

is satisfied as $T \to \infty$.

(iv) Let $\Theta_{R,k} = \bigcup_{u\in[0,1]} \Theta_{u,k}$ be compact and

$$\sup_{\theta_k \in \Theta_{R,k}} \|\Gamma_k^{-1}(\theta_k)\|_{sp} = O(k) \ , \quad \liminf_{T\to\infty} \int_0^1 [\Gamma_k^{-1}(\theta_{0,k}(u))]_{1,1} du \geq c > 0.$$

In order do illustrate the construction of the sets $\Theta_{u,k,i}$ in Assumption 4.1, consider exemplarily the case where for some $\delta > 0$ the polynomial $z \to 1 + \sum_{j=1}^{\infty} a_{j,0}(u) z^j$ with the coefficients from (2.7) is bounded away from zero inside the disc $D_\delta := \{z : |z| \leq 1 + \delta\}$ (uniformly with respect to u). In this case the sets $\Theta_{u,k,1} \times ... \times \Theta_{u,k,k}$ can be chosen as the intersection of the set $\{(\theta_{u,k,1}, \dots, \theta_{u,k,k}) \in \mathbb{R}^k || 1 + \sum_{j=1}^k \theta_{u,k,j} z^j| > C_1 > 0 \, \forall z \in D_\delta\}$ with the set

$$\{(a_1, ..., a_k) \in \mathbb{R}^k : \text{ there exists a sequence } (a_i)_{i>k} \text{ such that } (a_i)_{i\in\mathbb{N}} \in A_0\}.$$

Here the set A_0 is defined by

$$\begin{aligned} A_0 \quad := \quad & \Big\{(a_i)_{i\in\mathbb{N}} \Big| \text{ the function } p(z) := 1 + \sum_{j=1}^{\infty} a_j z^j \text{ satisfies } |p(z)| > C_2 > 0 \\ & \qquad \text{and } |p^{(l)}(z)| \leq C_3 \text{ for all } z \in D_\delta \text{ and } 0 \leq l \leq 4\Big\}, \end{aligned}$$

the constants C_2, C_3 are chosen such that $C_1 < C_2$ and such that the sequence $(a_{j,0})_{j\in\mathbb{N}}$ is an inner point of the set A_0.

Assumption (i) and (ii) are rather standard in a semi-parametric locally stationary time series model [see for example Dahlhaus and Giraitis (1998) or Dahlhaus and Polonik (2009) among others]. Note that the number of parameters k grows with increasing sample size in order to obtain a consistent estimate of the function $u \to d(u)$ in model (2.5). The restriction on the spectral norm in part (iv) was verified for a large number of long-range dependent models by Bhansali et al. (2006) [see equation (4.4) in this reference]. Note that these assumptions solely depend on the "true" underlying model.

On the other hand, an important step of our approach is the approximation of the spectral density $f(u, \lambda)$ in (2.6) by the truncated analogue

$$|1 - e^{i\lambda}|^{-2d_0(u)} |1 + \sum_{j=1}^{k} a_{j,0}(u) e^{-i\lambda j}|^{-2},$$

and the following assumption guarantees that the corresponding approximation error converges to 0 with reasonable rate. As a consequence it provides a link between the growth rate of $k = k(T)$ and N as the sample size T increases.

Assumption 4.2. *Suppose that $N \to \infty$, $N \log(N) = o(T)$ and*

$$\sup_{u\in[0,1]} \sum_{j=k+1}^{\infty} |a_{j,0}(u)| \quad = \quad O(N^{-1+\epsilon}) \tag{4.7}$$

for some $0 < \epsilon < 1/6$ as $T \to \infty$.

Note that

$$\begin{aligned} f(u, \lambda) - f_{\theta_{0,k}(u)}(\lambda) = {} & |1 - e^{i\lambda}|^{-2d_0(u)} \\ & \times \left(\big|1 + \sum_{j=1}^{\infty} a_{j,0}(u) e^{-i\lambda j}\big|^{-2} - \big|1 + \sum_{j=1}^{k} a_{j,0}(u) e^{-i\lambda j}\big|^{-2} \right), \end{aligned} \tag{4.8}$$

and an application of Lemma 2.4 in Kreiß et al. (2011) to the second factor (corresponding to the "short memory" part) shows that condition (4.7) with

$0 < \gamma_k < 1/2 - D$ implies

$$\sup_{u\in[0,1]} \int_{-\pi}^{\pi} |1-e^{i\lambda}|^{-2\gamma_k} \big| f(u,\lambda) - f_{\theta_{0,k}(u)}(\lambda) \big| d\lambda \quad = \quad O(N^{-1+\epsilon}).$$

As a consequence Assumption 4.1 (iii) is rather intuitive, because the parametric model (4.1) can be considered as an approximation of the "true" model defined in terms of the time varying spectral density (2.5). We finally note that condition (4.7) is satisfied for a large number of tvFARIMA(p,d,q) models, because it can be shown by similar arguments as in the proof of Theorem 2.2 in Preuß and Vetter (2013) that the coefficients $a_{j,0}(u)$ are geometrically decaying. This yields $\sum_{j=k+1}^{\infty} \sup_u |a_{j,0}(u)| = O(q^k)$ for some $q \in (0,1)$ resulting in a logarithmic growth rate for k, which is in line with the findings of Bhansali et al. (2006). Similarly, one can include processes whose AR coefficients decay such that $\sum_{j=0}^{\infty} \sup_u |a_{j,0}(u)| j^r < \infty$ is satisfied for some $r \in \mathbb{N}_0$. In this case k needs to grow at some specific polynomial rate.

Our first main result states a uniform convergence rate for the difference between $\hat{\theta}_{N,k}(u)$ and its true counterpart $\theta_{0,k}(u)$. As a consequence it implies that the estimator $\hat{d}_N$ obtained in the approximating models is uniformly consistent for the (time varying) long-range dependence parameter of the locally stationary process.

Theorem 4.3. *Let Assumption 2.1, 4.1 and 4.2 be satisfied and suppose that the estimator of the mean function μ satisfies*

$$N^{\epsilon} k^3 \max_{t=1,\dots,T} \big|\mu(t/T) - \hat{\mu}(t/T)\big| \quad = \quad o_p(1) \tag{4.9}$$

for some $0 < \epsilon < \min\{1/4 - D/2, 1/6\}$. If $N^{5/2}/T^2 \to 0$ and $k^4 \log^2(T) N^{-\epsilon/2} \to 0$, then

$$\begin{aligned} &\sup_{u\in[0,1]} \big\|\hat{\theta}_{N,k}(u) - \theta_{0,k}(u)\big\|_2 && (4.10) \\ &\quad = O_P\big(k^{3/2} N^{-1/2+\epsilon} + N^{\epsilon} k^{3/2} \max_{t=1,\dots,T} \big|\mu(t/T) - \hat{\mu}(t/T)\big|\big). \end{aligned}$$

In particular

$$\sup_{u\in[0,1]} |\hat{d}_N(u) - d_0(u)| = O_P\big(k^{3/2} N^{-1/2+\epsilon} + N^{\epsilon} k^{3/2} \max_{t=1,\dots,T} \big|\mu(t/T) - \hat{\mu}(t/T)\big|\big).$$

Remark 4.4. It follows from the proof of Theorem 4.10 below that there exists an estimator $\hat{\mu}$ with

$$N^{1/2-D-\alpha} \max_{t=1,\dots,T} \big|\mu(t/T) - \hat{\mu}(t/T)\big| = o_p(1)$$

for every $\alpha > 0$. Under the addional assumption

$$\sup_{u\in[0,1]} \sum_{j=k+1}^{\infty} |a_{j,0}(u)| \quad = \quad O(q^k) \tag{4.11}$$

for some $q \in (0,1)$ a logarithmic rate for the dimension k of the tvFARIMA models can be used such that assumption (4.9) is satisfied [for a broad class of models, where the stronger condition (4.11) is in fact satisfied, we refer to the discussion following (4.8)].

In order to obtain an estimator of the quantity F in (2.14) we assume without loss of generality that the sample size T can be decomposed into M blocks with length N (i.e. $T = NM$), where M is some positive integer. We define the corresponding midpoints in both the time and rescaled time domain by $t_j = N(j-1) + N/2$, $u_j = t_j/T$, respectively, and calculate $\hat{d}_N(u_j)$ on each of the M blocks as described in the previous paragraph. The test statistic is then obtained as

$$\hat{F}_T = \frac{1}{M}\sum_{j=1}^{M} \hat{d}_N(u_j). \tag{4.12}$$

The following two results specify the asymptotic behaviour of the statistic $\hat{F}_T$ under the null hypothesis and alternative.

Theorem 4.5. *Assume that the null hypothesis H_0 (of no long-range dependence) is true. Let Assumptions 2.1, 4.1 and 4.2 be satisfied, define $W_T = [\int_0^1 \Gamma_k^{-1}(\theta_{0,k}(u))du]_{1,1}$ and suppose that the estimator $\hat{\mu}$ of the mean function satisfies*

$$\max_{t=1,\dots,T} \big|\mu(t/T) - \hat{\mu}(t/T)\big| = O_p(N^{-1/2+\epsilon/2}), \tag{4.13}$$

$$\max_{t=1,\dots,T} \Big|\Big\{\mu\Big(\frac{t-1}{T}\Big) - \hat{\mu}\Big(\frac{t-1}{T}\Big)\Big\} - \Big\{\mu\Big(\frac{t}{T}\Big) - \hat{\mu}\Big(\frac{t}{T}\Big)\Big\}\Big| = O_p(N^{-1/2-2\epsilon}T^{-1/2}), \tag{4.14}$$

where ϵ is the constant in Assumption 4.2 satisfying $0 < \epsilon < 1/6$. Moreover, if the conditions

$$k^6\sqrt{T}/N^{1-\epsilon} \to 0, \quad k^4 \log^2(T)/N^{\epsilon/2} \to 0,$$
$$k^2 \log(T)/T^{1/6-\epsilon} \to 0, \quad k^2N^2/T^{\frac{3}{2}} \to 0$$

hold as $T \to \infty$, then we have

$$\sqrt{T}\hat{F}_T/\sqrt{W_T} \xrightarrow{D} \mathcal{N}(0,1). \tag{4.15}$$

Note that $\hat{F}_T$ is an average of the estimates of the long-range dependence parameter in the approximating tvFARIMA model. By Assumption 4.1 the point 0 is an interior point of the canonical projection of the parameter space $\Theta_{u,k}$ onto the first component, which motivates the asymptotic normality obtained in Theorem 4.5. More precisely, we show in Section 7 that the leading term in the stochastic expansion of $\hat{F}_T$ is given by

$$-\frac{1}{M}\sum_{j=1}^{M}\frac{1}{4\pi}\int_{-\pi}^{\pi}\big(I_N^{\mu}(u_j,\lambda) - f_{\theta_{0,k}(u_j)}(\lambda)\big)[\Gamma_k^{-1}(\theta_{0,k}(u_j))\nabla f^{-1}_{\theta_{0,k}(u_j)}(\lambda)]_1\, d\lambda,$$

where $[a]_1$ denotes the first element of the $(k+1)$ dimensional vector a. Asymptotic normality follows because the individual terms in this sum are asymptotically independent (see Section 7 for details and Section 3 for a similar result in a simplified model).

Theorem 4.6. *Assume that the alternative H_1 of long-range dependence is true. Let Assumptions 2.1, 4.1 and 4.2 be satisfied and suppose that the estimator $\hat{\mu}$ of the mean function satisfies*

$$N^{\epsilon} k^3 \max_{t=1,\ldots,T} \left|\mu(t/T) - \hat{\mu}(t/T)\right| \quad = \quad o_p(1), \tag{4.16}$$

where ϵ is the constant in Assumption 4.2 satisfying $0 < \epsilon < \min\{1/4 - D/2, 1/6\}$. Moreover, if the conditions

$$k^6/N^{1-2\epsilon} \to 0, \quad k^4 \log^2(T)/N^{\epsilon/2} \to 0, \quad k^4/N^{1-2D-2\epsilon} \to 0, \quad k^2 N^{5/2}/T^2 \to 0$$

are satisfied as $T \to \infty$, then we have

$$\hat{F}_T \xrightarrow{P} F > 0.$$

Remark 4.7. (*more transparent conditions*) If assumption (4.11) is satisfied, more transparent conditions for Theorem 4.3, 4.5 and 4.6 can be given. To be precise assume that (4.11) holds for some $q \in (0,1)$ and choose

$$k = \lfloor -a \frac{\log T}{\log q} \rfloor$$

for some $a \in (1/2, 1)$. If $D < 1/6$, then it follows by straightforward but tedious calculations that Theorem 4.3, 4.5 and 4.6 hold for $N = T^{\beta}$ with any β satisfying $a < \beta < \min\{\frac{6}{5}a, \frac{3}{4}\}$ (note that this conditions provides a further restriction for the choice of the constant a). Similarly, if $1/6 \le D < 1/2$ the results hold, whenever $a < \beta < \min\{\frac{4a}{3+2D}, \frac{3}{4}\}$.

Remark 4.8. (*the non-Gaussian case*) It is worthwhile to mention that in most of articles cited in this paper the assumption of Gaussianity for the innovation process in (2.1) is required. In the present case this assumption is not necessary and is only imposed here to simplify technical arguments in the proof of Theorem 7.1. This observation is a consequence of method of proof used in Section 7. In fact, asymptotic normality is established by the method of moments showing that all cumulants of the statistic under consideration converge to those of a normal distribution. In the definition of all cumulants one needs the existence of all moments of Z_i (which is obviously true in the Gaussian case). The main simplification under the assumption of Gaussianity consists in the fact that one does not have to work with partitions including cumulants of any possible order. The extension to non Gaussian innovations does not change the main argument in the proofs, but the calculations become substantially more complicated, and the details are omitted for the sake of brevity.

As a consequence all results of this section remain true as long as the innovations are independent with all moments existing, mean zero and $\mathbb{E}(Z_t^2) =$

$\sigma^2(t/T)$ for some twice continuously differentiable function $\sigma : [0,1] \to \mathbb{R}$. To be more precise, in order to address for non-Gaussian innovations the variance V_T in Theorem 7.1 (which is one of the main ingredients for the proofs in Section 7) has to be replaced by

$$V_{T,general} = V_T + \frac{1}{TM}\sum_{j=1}^{M}\kappa_4(u_j)/\sigma^4(u_j)\Big(\int_{-\pi}^{\pi} f(u_j,\lambda)\phi_T(u_j,\lambda)d\lambda\Big)^2,$$

where V_T is defined in (7.5) and $\kappa_4(u)$ denotes the fourth cumulant of the innovations, i.e. $\kappa(t/T) = \mathbb{E}(Z_t^4) - 3\sigma^4(t/T)$ for all $t = 1, \dots, T$. In the proof of Theorem 4.5 we apply this result with $\phi_T(u_j,\lambda) = (4\pi)^{-1}[\Gamma_k^{-1}(\theta_{0,k}(u_j))\nabla f^{-1}_{\theta_{0,k}(u_j)}(\lambda)]_1$. Consequently, we obtain that in the non-Gaussian case the asymptotic normality in Theorem 4.5 holds, where the matrix $W_T = [\int_0^1 \Gamma_k^{-1}(\theta_{0,k}(u))du]_{1,1}$ has to be replaced by

$$W_{T,general} = W_T + \frac{1}{TM}\sum_{j=1}^{M}\kappa_4(u_j)/\sigma^4(u_j)\Big(\int_{-\pi}^{\pi} f(u_j,\lambda)\phi_T(u_j,\lambda)d\lambda\Big)^2. \quad (4.17)$$

Thus, under the null hypothesis it follows that

$$\frac{\sqrt{T}\hat{F}_T}{\sqrt{W_{T,general}}} \xrightarrow{D} \mathcal{N}(0,1). \quad (4.18)$$

Remark 4.9. (*the final test*) Note that the first term W_T in (4.17) can be consistently estimated by

$$\hat{W}_T = \Big[\frac{1}{M}\sum_{j=1}^{M}\Gamma_k^{-1}(\hat{\theta}_{N,k}(u_j))\Big]_{11}.$$

This gives as an estimator for $V_{T,\mathrm{general}}$ the statistic

$$\hat{W}_{T,general} = \hat{W}_T + \frac{1}{M}\sum_{j=1}^{M}\frac{\hat{\kappa}_4(u_j)}{\hat{\sigma}^4(u_j)}\Big(\int_{-\pi}^{\pi} f_{\hat{\theta}_{N,k}(u_j)}(\lambda)[\Gamma_k^{-1}(\hat{\theta}_{N,k}(u_j))\nabla f^{-1}_{\hat{\theta}_{N,k}(u_j)}(\lambda)]_1 d\lambda\Big)^2,$$

where $\hat{\sigma}(u_j)$ and $\hat{\kappa}(u_j)$ are obtained by calculating the empirical second and fourth moment $\hat{\mu}_{2,Z}(u_j)$, $\hat{\mu}_{4,Z}(u_j)$ of the residuals

$$Z_{t,res} = X_{t,T} - \sum_{i=2}^{k}[\hat{\theta}_{N,k}(u_j)]_i X_{t-i+1,T},$$
$$t = t_j - N/2 + k + 1, t_j - N/2 + k + 2, ..., t_j + N/2,$$

and setting $\hat{\sigma}^2(u_j) = \hat{\mu}_{2,Z}(u_j)$, $\hat{\kappa}(u_j) = \hat{\mu}_{4,Z}(u_j) - 3\hat{\mu}^2_{2,Z}(u_j)$. Since

$$\hat{W}_{T,general}/W_{T,general} \xrightarrow{P} 1,$$

an asymptotic level α-test is obtained from (4.18) by rejecting the null hypothesis (2.13), whenever

$$\sqrt{T}\hat{F}_T/\sqrt{\hat{W}_{T,general}} \geq u_{1-\alpha}, \tag{4.19}$$

where $u_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the standard normal distribution (in the Gaussian case $\hat{W}_{T,general}$ can be replaced by $\hat{W}_T$). It then follows from Remark 4.8 and Theorem 4.6 that for any estimator of the mean function μ satisfying (4.13), (4.14) and (4.16), the test, which rejects H_0 whenever (4.19) is satisfied, is a consistent level-α test for the null hypothesis stated in (2.13). The finite sample properties of this resulting test are investigated in Section 5.

A popular estimate of the mean function is given by the the local-window estimator

$$\hat{\mu}_L(u) = \frac{1}{L}\sum_{p=0}^{L-1} X_{\lfloor uT \rfloor - L/2+1+p,T} , \tag{4.20}$$

where L is a window-length which does not necessarily coincide with the corresponding parameter in the calculation of the local periodogram. Note that also $I_N^{\hat{\mu}}(u,\lambda)$ is an asymptotically unbiased estimator for $f(u,\lambda)$ if $N \to \infty$ and $N/T \to 0$. The final result of this section shows that this estimator satisfies the assumptions of Theorem 4.5 and 4.6 if L grows at a 'slightly' faster rate than N. This means, it can be used in the asymptotic level α test defined by (4.19).

Theorem 4.10. *a) Suppose that the assumptions of Theorem 4.5 hold and additionally $N^{1+4\epsilon}/L^{1-\delta} \to 0$ and $L^{5/2-\delta}/T^{3/2} \to 0$ are satisfied for some $\delta > 0$, where $\epsilon > 0$ denotes the constant in Theorem 4.5. Then the local-window estimator $\hat{\mu}_L$ defined in (4.20) satisfies (4.13) and (4.14).*

b) Suppose the assumptions of Theorem 4.6 hold. If additionally $N^\epsilon k^5/L^{1/2-D-\delta} \to 0$ and $L^{5/2-D}/T^2 \to 0$ for some $0 < \delta < 1/2 - D - \epsilon$ (with the constant ϵ from Theorem 4.6), then the local-window estimator $\hat{\mu}_L$ defined in (4.20) satisfies (4.16).

Remark 4.11. (*parametric models*) Analogues of Theorem 4.5 and 4.6 can be obtained in a parametric framework. To be precise, assume that the approximating processes $\{X_t(u)\}_{t\in\mathbb{Z}}$ has a time varying spectral density of the form (4.1), where k is fixed and known. In this case it is not necessary that the dimension k is increasing with the sample size T and Assumption 4.1(iii) and 4.2 are not required. All other stated assumptions are rather standard in this framework of a semi-parametric locally stationary time series model [see for example Dahlhaus and Giraitis (1998) or Dahlhaus and Polonik (2009) among others]. With these modifications Theorem 4.5 and 4.6 remain valid and as a consequence we obtain an alternative test to the likelihood ratio test proposed in Yau and Davis (2012), which operates in the spectral domain and can be used for more general null hypotheses as considered by these authors.

Remark 4.12. (*local alternatives*) Theorem 4.5 remains valid under local alternatives converging to the null hypothesis at a rate $\sqrt{T/k}$. To be precise let $d_{0,T}(u) = a(u)\sqrt{W_{T,\text{general}}/T}$ where $a : [0,1] \to [0,\infty)$ is a twice continuously differentiable function such that $\int_0^1 a(u)\,du > 0$. Then it follows by similar arguments as given in the proof of Theorem 4.5, that

$$\sqrt{T}\Big(\frac{\hat{F}_T - \int_0^1 a(u)\,du}{\sqrt{W_{T,\text{general}}}}\Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

(note that $W_{T,\text{general}} = O(k)$ due to Assumption 4.1(iv) and that W_T does not depend on the long-memory parameter function d_0). This indicates that (asymptotically) the power of the test (4.19) is increasing with $\int_0^1 a(u)\,du$, which can also be observed in the simulation study presented in the following Section.

Remark 4.13. (*two sided hypotheses*) In some cases it might also be of interest to detect any type of "fractional" behaviour either with positive or negative values of $d_0(u)$ depending on the period. In principle, problems of this type can be addressed by rejecting the null hypothesis

$$H_0^{(1)} : \int_0^1 |d_0(u)| du = 0$$

for large values of the statistic

$$\hat{F}_T^{(1)} = \frac{1}{M} \sum_{j=1}^{M} |\hat{d}_N(u_j)|.$$

The analysis of the asymptotic properties of this statistic is not a direct consequence of the arguments presented in Section 7. Observing Section 3.4 we conjecture that an appropriately standardized version of $\hat{F}_T^{(1)}$ is also asymptotically normal distributed, but a rigorous proof would be beyond the scope of the present paper.

Remark 4.14. (*alternative parameter estimates)* The parameters of the approximating time varying FARIMA process are estimated by a local Whittle estimator from an approximating time varying FARIMA process, but other estimators for the time varying long range dependence parameter could be used as well. Alternative estimators are localized versions of the log-periodogram and the "local" Whittle estimator who have been considered and investigated in the stationary case by Geweke and Porter-Hudak (1983); Robinson (1995) and Künsch (1987); Robinson (1994), respectively. We expect that the use of estimators of this type yields to similar results as stated in Theorem 4.5 and 4.6, but the asymptotic variance in Theorem 4.5 will probably be different. As these estimators have been designed to avoid strong smoothness assumptions regarding the spectral density, their use in the procedure proposed in this paper may lead to weaker assumptions in Theorem 4.3 and 4.5. For a rigorous proof of results of this type one has to extend the results Robinson (1994) and Robinson (1995) to a time varying long range dependence parameter (including a definition of appropriate estimators for this case) and then carefully modify the arguments given in Section 7.2 and and 7.3.

In order to indicate the difficulties which might occur in the use of different estiamtors, we consider a further alternative estimator, which is - on a first glance - very closely related to the estimator investigated in this section. Note, that from a computational point of view, it might be simpler to replace the integral in (4.3) by a sum over Fourier frequencies. Again we expect that similar results as presented in this section could be obtained for this alternative estimator, but all arguments in the proofs in Section 7 have to be checked very carefully and to modified at several places. In particular the standardization in (4.15) may change if integrals are replaced by sums over Fourier frequencies. For example, the asymptotic variances of the statistics

$$\begin{aligned}\tilde{F}_{1,T} &= \frac{1}{4\pi M}\sum_{j=1}^{M}\int_{-\pi}^{\pi} I_N^2(u_j,\lambda)\,d\lambda,\\ \hat{F}_{1,T} &= \frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{\lfloor \frac{N}{2}\rfloor} I_N^2(u_j,\lambda_{k,N}),\end{aligned}$$

are different (here $\lambda_{k,N} =$ denotes the kth Fourier frequency) and given by

$$\lim_{T\to\infty} T\,Var\,(\tilde{F}_{1,T}) = \frac{14}{3\pi}\int_{-\pi}^{\pi}\int_0^1 f^4(u,\lambda)dud\lambda,$$
$$\lim_{T\to\infty} T\,Var\,(\hat{F}_{1,T}) = \frac{5}{\pi}\int_{-\pi}^{\pi}\int_0^1 f^4(u,\lambda)dud\lambda,$$

respectively [see Deo and Chen (2000) and Sen et al. (2016), who proved this result in the stationary short memory and locally stationary long memory case, respectively]. The answer to the question if a similar phenomenon occurs if a discretized Whittle estimator is used in the test statistic (4.12) is by no means trivial and requires a new proof following the arguments outlined in Section 7.

Remark 4.15. (*weaker assumptions and generalizations*)
(1) The (uniform) smoothness conditions stated in Assumption 2.1 are commonly made in the literature [see for example Palma and Olea (2010)] and are also required in the present context to obtain the uniform consistency of the estimator for the function d_0. However, it is worthwhile to mention that the asymptotic properties of the proposed test can also be derived under weaker assumptions. To be more precise, Theorem 4.5 remains valid if the conditions on the function $\psi_l(u)$ and its derivatives stated in Assumption 2.1 are replaced by

$$\begin{aligned}|\psi_l'(u)| &\leq C(u)\log|l||l|^{D-1}, |\psi_l''(u)| \leq C(u)\log^2|l||l|^{D-1},\\ \Big|\frac{\partial}{\partial u}f(u,\lambda)\Big| &\leq C(u)|\log(\lambda)||\lambda|^{-2D}, \Big|\frac{\partial^2}{\partial u^2}f(u,\lambda)\Big| \leq C(u)\log^2(\lambda)|\lambda|^{-2D}\end{aligned}$$

for all $\lambda \in [-\pi,\pi]$ and $u \in (0,1)$. Here $C : (0,1) \to \mathbb{R}$ denotes a function such that $\int_0^1 |C(u)|^p du < \infty$ for all $p \in \mathbb{N}$. The proof of this statement can be performed by similar arguments as given in the proof of Theorem 4.5 with additional technical arguments for the more delicate estimates of the error terms.

Moreover, we conjecture that, the conditions can be further weakened such that the function C is only integrable up to a specific order. A detailed verification of such a statement, however, is an open problem and far beyond the scope of the present paper.

(2) As pointed out by a referee a further important question in this context is to extend the theory to mean functions with jumps. For this purpose one would have to construct uniformly consistent estimators of a piecewise continuous mean function (such that assumptions (4.9) or (4.13) are satisfied). However, this is a very difficult task and we are not aware of any estimators with this property. Even in the classical case of nonparametric regression with independent errors, there are no results available regarding the uniform consistency of an estimate of the regression functions in the case of jumps (to our best knowledge). Recent papers provide the uniform consistency over sets which exclude the jump points [see Gijbels et al. (2007) and Xia and Qiu (2012) among others]. However, these results do not yield a uniform convergence of the type (4.9) or (4.13) required in our theory. Moreover, is also not clear if the results of these authors can be generalized for locally stationary long range dependent time series as considered in this paper.

(3) A further interesting problem for future research is the extension of the results to processes with a long range dependence parameter $d_0(u) > 1/2$. While this seems to be difficult for the Whittle estimator (4.3) considered in this paper, it might be possible for the semiparametric estimators investigated by Robinson (1994, 1995), who considered stationary models with a spectral density satisfying $f(\lambda) \approx G|\lambda|^{-2d}$ as $\lambda \to 0$, where $d \in (-1/2, 1/2)$. There exist several references investigating asymptotic properties of the local Whittle estimator introduced by Künsch (1987) in the case of fractional processes, that is $d > 1/2$ [see Velasco (1999) or Phillips and Shimotsu (2004) among others]. A very challenging question is, if similar results as derived in this section can be obtained for models with a *time varying* spectral density satisfying

$$f(\lambda, u) \approx |\lambda|^{-2d_0(u)} g(u, \lambda) \quad \text{as} \quad \lambda \to 0 \; .$$

In the simplest case, where the function g does not depend on the frequency, that is $g(u, \lambda) = g(u)$, one can use a localized version of the local Whittle estimator proposed by Künsch (1987) to estimate the function d_0, but the theoretical properties of such a statistic are not obvious. In particular it is not clear if this approach yields to a uniformly consistent estimate (including the appropriate rates) as derived in Theorem 4.3 and required for the theory presented in this paper.

5. Finite sample properties

The application of the test (4.19) requires the choice of several parameters. Based on an extensive numerical investigation we recommend the following rules.

For the choice of the parameter L in the local window estimate $\hat{\mu}_L$ of the mean function [for a precise definition see (4.20)] we use $L = N^{1.05}$. Because the procedure is based on a sequence of approximating tvFARIMA$(k, d, 0)$-processes the choice of the order k is essential, and we suggest the AIC criterion for this purpose, that is

$$\hat{k} = \arg\min_k \frac{1}{T} \sum_{j=1}^{T/2} \Big(\log(h_{\hat{\theta}_{k,s}}(\lambda_j)) + \frac{I^{\hat{\mu}}(\lambda_j)}{h_{\hat{\theta}_{k,s}}(\lambda_j)} \Big) + \frac{k+1}{T}, \tag{5.1}$$

where $\lambda_j = 2\pi j/T$ $(j = 1, \dots, T)$, and $h_{\hat{\theta}_{k,s}}(\lambda)$ is the estimated spectral density of a stationary FARIMA$(k, d, 0)$ process and $I^{\hat{\mu}_L}(\lambda)$ is the mean-corrected periodogram given by

$$I^{\hat{\mu}_L}(\lambda) \quad := \quad \Big| \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^{T} \Big[X_{t,T} - \hat{\mu}_L(t/T) \Big] e^{-it\lambda} \Big|^2.$$

Note that we choose the same order k for each of the M blocks. An alternative choice is to use tvFARIMA models of different order for each block. In our numerical experiments we investigated both methods and we observed substantial advantages for the rule (5.1) (the results of this comparison are not displayed for the sake of brevity). Because this approach also has additional computational advantages we recommend to choose the same approximating tvFARIMA(k,d,0) model for all blocks. Finally, the performance of the test depends on the choice of N, and this dependency will be carefully investigated in the following discussion.

5.1. Simulation of level and power

All results presented in this section are based on 1000 simulation runs, and we begin with an investigation of the approximation of the nominal level of the test (4.19) considering three examples. The first example is given by a location model with a tvAR(1)-process, that is

$$X_{t,T} \quad = \quad \mu_i(t/T) + Y_{t,T}, \quad t = 1, \dots, T, \tag{5.2}$$

where

$$Y_{t,T} = 0.6 \frac{t}{T} Y_{t-1,T} + Z_t, \quad t = 1, \dots, T. \tag{5.3}$$

The innovations $\{Z_t\}_{t=1,\dots,T}$ in (5.3) are either i.i.d. standard normal or i.i.d. chi-square distributed normalized such that $E[Z_i] = 0, Var(Z_i) = 1$, i.e. $Z_i \sim (\chi^2_5 - 5)/\sqrt{10}$. Two cases are investigated for the mean function representing a smooth change and abrupt change in the mean effect, i.e.

$$\mu_1(t/T) \quad = \quad 1.2 \frac{t}{T}, \tag{5.4}$$

$$\mu_2(t/T) \quad = \quad \begin{cases} 0.65 & \text{for } t = 1, \dots, T/2 \\ 1.3 & \text{for } t = T/2+1, \dots T. \end{cases} \tag{5.5}$$

The mean function (5.5) is not smooth and used to investigate the impact of a violation of the assumptions in the procedure. Our third example consists of a tvMA(1)-process given by

$$X_{t,T} = Z_t + 0.55 \sin\left(\pi \frac{t}{T}\right) Z_{t-1}, \quad t = 1, \dots, T, \tag{5.6}$$

where $\{Z_t\}_{t=1,\dots,T}$ is again a sequence of i.i.d. normal or chi-square distributed random variables normalized to have mean 0 and variance 1. Figure 3 and 4 show the sample autocovariance and the sample partial autocovariance functions of 1024 observations generated by the models (5.4), (5.5) and (5.6), respectively, from which it is clearly visible that the mean functions in (5.4) and (5.5) are causing a long-memory type behaviour. In Table 3, we show for these models the simulated level of the test (4.19) for various choices of N. We observe in model (5.2) and (5.6) a reasonable approximation of the nominal level whenever $M = T/N \approx 4$ and the sample size T is larger or equal than 512. Here the results are similar for normal and chi-square distributed innovations. On the other hand in model (5.2) with mean function (5.5) the assumptions of the asymptotic theory are violated and the situation is different. For moderate sample sizes the specification $M = T/N \approx 4$ yields to an overestimation of the nominal level. Moreover, the approximation of the nominal level becomes worse with increasing sample size. We conjecture that the performance of the test could be improved by using estimators addressing the problem of jumps in the mean function.

In order to investigate the power of the test (4.19) and to compare it with the competing procedures proposed by Berkes et al. (2006), Baek and Pipiras (2012) and Yau and Davis (2012), we simulated data from a tvFARIMA$(1, d, 0)$-process

$$(1 + 0.2\frac{t}{T}B)(1 - B)^{d(t/T)} X_{t,T} = Z_t, \quad t = 1, \dots, T, \tag{5.7}$$

and a tvFARIMA$(0, d, 1)$-process

$$(1 - B)^{d(t/T)} X_{t,T} = (1 - 0.35\frac{t}{T}B) Z_t, \quad t = 1, \dots, T, \tag{5.8}$$

where B is the backshift operator that is $B^j X_{t,T} := X_{t-j,T}$. In both cases the long-memory function is given by $d(t/T) = 0.1 + 0.3t/T$. Because all competing procedures are designed to detect stationary long-range dependent alternatives, we also simulated data from a stationary FARIMA(1,d,1)-process

$$(1 + 0.25B)(1 - B)^{0.1} X_T = (1 - 0.3B) Z_t, \quad t = 1, \dots, T. \tag{5.9}$$

The corresponding results for the new test (4.19) and its competitors are presented in Table 4-7. In Table 4 and 5 we show the simulated power in model (5.7) for (standardized) normal and chi-square distributed innovations. We do not observe substantial differences in the power of the new test under different distributional assumptions and for this reason Table 6 and 7 only contain results

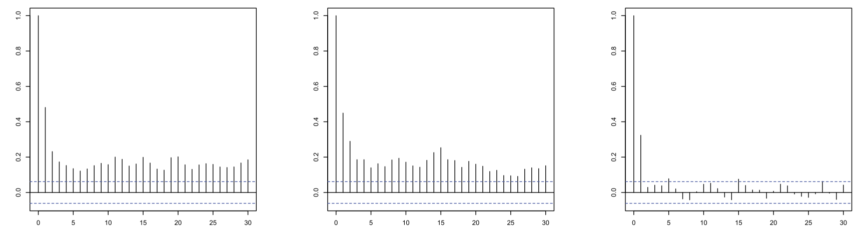

FIG 3. *Sample autocovariance functions of model (5.2) with mean function (5.4) (left panel), (5.5) (middle panel) and of model (5.6) (right panel). The sample size is T=1024.*

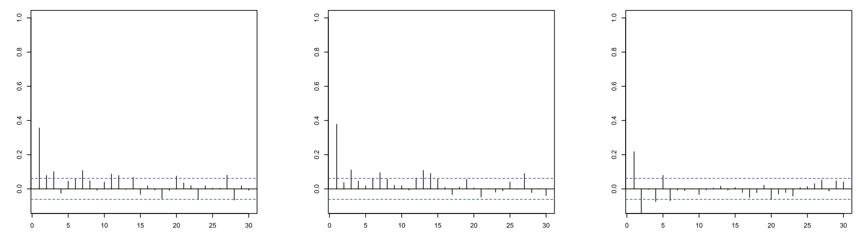

FIG 4. *Sample partial autocovariance functions of model (5.2) with mean function (5.4) (left panel), (5.5) (middle panel) and of model (5.6) (right panel). The sample size is T=1024.*

for normal distributed innovations. In the first column the rejection probabilities of the new test are displayed and we observe a reasonable power in all models under consideration. Interestingly, the differences in power between the tvFARIMA$(1, d, 0)$ and the tvFARIMA$(0, d, 1)$-model are rather small (see second column in Table 4 and 6). The results in Table 7 show a loss in power, which corresponds to intuition because the "average" long-memory effect in model (5.9) is 0.1, while it is $\int_0^1 (0.1 + 0.3u)\, du = 0.25$ in model (5.7) and (5.8) [see also Remark 4.12 and the discussion at the end of this section]. In order to compare the new test with existing approaches we next investigate the performance of the procedures proposed by Berkes et al. (2006), Baek and Pipiras (2012) and Yau and Davis (2012), which are designed for a test of the null hypothesis "the process has the short memory property with a structural break in the mean" against the alternative "the process is stationary and has the long memory property". The third columns of Table 4-7 show the power of the test in Baek and Pipiras (2012), which also operates in the spectral domain. These authors estimate the change in the mean with a break point estimator and remove this mean effect (which is responsible for the observed local stationarity) from the time series. Then they calculated the local Whittle estimator introduced by Robinson (1995) for the self similarity parameter and reject the null hypothesis for large value of this estimate. Note that the calculation of the local Whittle estimator requires the specification of the number of "low frequencies" and we used $m = \sqrt{T}$ as Baek and Pipiras (2012) suggested in their simulation study. We observe that the new test (4.19) yields larger power than the procedure of Baek and Pipiras (2012) in nearly all cases under consideration. This improvement becomes more substantial with increasing sample size.

Next we study the performance of the procedure proposed by Berkes et al. (2006) in models (5.7)-(5.9). These authors use a CUSUM statistic to construct an estimator, say $\hat{k}^*$, for a (possible) change point k^* in a time series. Then two CUSUM statistics are computed for the first $\hat{k}^*$ elements of the time series and the remaining ones, respectively. The test statistic is given by the maximum of those two. For the choice of the bandwidth function we use $q(n) = 15\log_{10}(n)$ as suggested by these authors in Section 3 of their article. The results are depicted in the fourth columns of Table 4-7 and demonstrate that this test is not able to detect long-range dependence in both the stationary and locally stationary case. These findings coincide with the results of Baek and Pipiras (2012) who also remarked that the test in Berkes et al. (2006) has very low power against long-range dependence alternatives.

The method proposed by Yau and Davis (2012) consists of a parametric likelihood ratio test assuming two (not necessarily equal) ARMA(p, q) models before and after the breakpoint of the mean function. Their method requires a specification of the order of these two models and we used ARMA$(1, 1)$-models under the null hypothesis and a FARIMA$(1, d, 1)$ model under the alternative hypothesis. The corresponding results for this test are depicted in the fifth columns of Table 4-7 corresponding to non-stationary and stationary long-range dependent alternatives, respectively. We observe that in these cases the new test (4.19) outperforms the test proposed in Yau and Davis (2012) if the sample size is larger than 512 and that both tests have similar power for sample size 256 (see the fifth column of Table 4 and 6). On the other hand, in the case of the long-range dependent stationary alternative (5.9) the test of Yau and Davis (2012) yields slightly better rejection probabilities than the new test (4.19) for smaller sample sizes while we observe advantages of the proposed test in this paper for sample sizes 512 and 1024. These results are remarkable, because the test of Yau and Davis (2012) is especially designed to detect stationary alternatives of FARIMA$(1, d, 1)$ type, but the new semi-parametric test still yields an improvement in many cases.

Finally, as it was pointed out by a reviewer, it is also of interest to systematically investigate the power of the test (4.19) as a function of the quantity $F = \int_0^1 d(u)\,du$. The arguments in Remark 4.12 indicate that the power is increasing with F, and we will now investigate if these properties can also be observed in finite samples. For this purpose we simulated data from the tvFARIMA(0,d,1)-process in (5.8) with different choices for the long-memory function d:

$$d_1(t/T) \;=\; 1/8 \quad , \quad d_2(t/T) = t/4T, \tag{5.10}$$

$$d_3(t/T) \;=\; \begin{cases} 0 & \text{for } 1 \le t \le T/3 \\ 3/8 & \text{for } T/3 < t \le 2T/3 \\ 0 & \text{for } 2T/3 < t \le T, \end{cases} \tag{5.11}$$

$$d_4(t/T) \;=\; 0.3 \quad , \quad d_5(t/T) = 1.8t/T(1 - t/T). \tag{5.12}$$

			$Z_t \sim \mathcal{N}(0,1)$					
			(5.2), (5.4)		(5.2),(5.5)		(5.6)	
T	N	M	5%	10%	5%	10%	5%	10%
256	64	4	.090	.128	.094	.145	.085	.122
256	32	8	.151	.228	.165	.255	.182	.261
512	128	4	.061	.095	.070	.114	.069	.099
512	64	8	.089	.130	.089	.126	.081	.107
1024	256	4	.046	.072	.077	.119	.069	.106
1024	128	8	.059	.087	.061	.088	.064	.093
2048	512	4	.048	.090	.094	.148	.074	.122
2048	256	8	.026	.034	.026	.058	.062	.084
4096	1024	4	.056	.094	.164	.248	.076	.112
4096	512	8	.014	.030	.026	.056	.060	.080
			$Z_t \sim (\chi^2_5 - 5)/\sqrt{10}$					
			(5.2), (5.4)		(5.2),(5.5)		(5.6)	
T	N	M	5%	10%	5%	10%	5%	10%
256	64	4	.094	.142	.100	.162	084	.118
256	32	8	.218	.319	.249	.335	.187	.258
512	128	4	.066	.100	.062	.098	.068	.090
512	64	8	.086	.144	.102	.140	.074	.118
1024	256	4	.042	.076	.080	.126	.080	.114
1024	128	8	.058	.082	.090	.124	.066	.106
2048	512	4	.048	.078	.116	.154	.086	.116
2048	256	8	.020	.026	.040	.068	.046	.074
4096	1024	4	.052	.098	.196	.264	.085	.127
4096	512	8	.026	.044	.046	.062	.058	.090

TABLE 3

Simulated level of the test (4.19) for different processes and choices of T,N and M.

			(4.19)		Baek/Pipiras		Berkes et. al		Yau/Davis	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%
256	64	4	0.288	0.354	0.248	0.330	0.037	0.080	0.250	0.306
256	32	8	0.290	0.436						
512	128	4	0.530	0.590	0.356	0.468	0.006	0.041	0.182	0.226
512	64	8	0.348	0.458						
1024	256	4	0.746	0.770	0.562	0.656	0.026	0.102	0.204	0.267
1024	128	8	0.412	0.512						
2048	512	4	0.882	0.900	0.724	0.816	0.152	0.222	0.376	0.452
2048	256	8	0.625	0.683						
4096	1024	4	0.974	0.978	0.892	0.928	0.318	0.460	0.740	0.782
4096	512	8	0.892	0.910						

TABLE 4

Rejection frequencies of the test (4.19) and three competing procedures under the alternative (5.7) for different choices of T,N and M. The innovations are standard normal distributed.

For the functions d_1, d_2, and d_3 the quantity $F = \int_0^1 d(u)\,du$ is given by 1/8 while $F = 3/10$ for d_4 and d_5. The corresponding results are shown in Table 8. We mainly discuss the case $M = 4$ (because it yields to the best approximation of the nominal level) and mention that the interpretation of the results for other choice of M is very similar. For a fixed $F = 1/8$ we do not observe substantial differences between the functions d_1 and d_2 in the case $M = 4$, while the function d_3 yields to a larger power. This observation can be explained by the fact that the integral in (2.14) is approximated by a Riemann sum $\frac{1}{M}\sum_{j=1}^M d(u_j)$ at points $u_j = \frac{j-1}{M} + \frac{1}{2M}$. Consider exemplarily the case $M = 4$ (which is recommended, because it yields to a good approximation of the nominal level). While for the function $d_1(u) = 1/8$ all estimates roughly yield the same contribution of size

			(4.19)		Baek/Pipiras		Berkes et. al		Yau/Davis	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%
256	64	4	0.340	0.436	0.244	0.343	0.034	0.082	0.262	0.335
256	32	8	0.373	0.492						
512	128	4	0.550	0.600	0.434	0.510	0.005	0.021	0.228	0.276
512	64	8	0.362	0.476						
1024	256	4	0.714	0.756	0.527	0.641	0.047	0.130	0.197	0.240
1024	128	8	0.446	0.522						
2048	512	4	0.910	0.926	0.721	0.805	0.143	0.244	0.263	0.334
2048	256	8	0.634	0.708						
4096	1024	4	0.974	0.976	0.889	0.938	0.311	0.408	0.713	0.741
4096	512	8	0.923	0.938						

Table 5

Rejection frequencies of the test (4.19) *and three competing procedures under the alternative (5.7) for different choices of T,N and M. The innovations are* $(\chi_5^2 - 5)/\sqrt{10}$ *distributed.*

			(4.19)		Baek/Pipiras		Berkes et. al		Yau/Davis	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%
256	64	4	0.260	0.330	0.230	0.322	0.039	0.088	0.296	0.366
256	32	8	0.276	0.394						
512	128	4	0.528	0.590	0.342	0.456	0.010	0.036	0.268	0.322
512	64	8	0.314	0.414						
1024	256	4	0.774	0.796	0.546	0.656	0.024	0.086	0.228	0.292
1024	128	8	0.414	0.492						
2048	512	4	0.900	0.913	0.758	0.820	0.168	0.268	0.320	0.404
2048	256	8	0.608	0.665						
4096	1024	4	0.994	0.996	0.900	0.940	0.332	0.444	0.649	0.697
4096	512	8	0.982	0.990						

Table 6

Rejection frequencies of the test (4.19) *and three competing procedures under the alternative (5.8) for different choices of T,N and M. The innovations are standard normal distributed.*

			(4.19)		Baek/Pipiras		Berkes et. al		Yau/Davis	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%
256	64	4	0.094	0.136	0.087	0.149	0.045	0.093	0.178	0.210
256	32	8	0.138	0.216						
512	128	4	0.146	0.196	0.119	0.177	0.022	0.055	0.140	0.176
512	64	8	0.138	0.214						
1024	256	4	0.328	0.406	0.127	0.197	0.018	0.079	0.152	0.206
1024	128	8	0.152	0.218						
2048	512	4	0.646	0.710	0.174	0.266	0.052	0.116	0.374	0.470
2048	256	8	0.312	0.388						
4096	1024	4	0.854	0.888	0.232	0.466	0.064	0.162	0.736	0.792
4096	512	8	0.716	0.742						

Table 7

Rejection frequencies of the test (4.19) *and three competing procedures under the alternative (5.9) for different choices of T,N and M. The innovations are standard normal distributed.*

1/8, we observe that for the function d_3 two points (namely u_2 and u_3) yield a contribution of size 3/8 and the other points u_1, u_4 yield the value $d_3(u_j) = 0$ ($j = 1, 4$). Nevertheless the total contribution in this case is 3/16, while it is only 1/8 for d_1. This explains the improvement in power observed for the function d_3. We expect that these advantages vanish asymptotically, because the approximation of $F = \int_0^1 d(u)\,du$ by its Riemann sum becomes more accurate with increasing M. Finally, a comparison of columns 1-3 (corresponding to the case $F = 1/8$ with columns 4-5 in Table 8 (corresponding to the case $d = 3/10$)

			d_1		d_2		d_3		d_4		d_5	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
256	64	4	0.118	0.174	0.118	0.168	0.146	0.222	0.374	0.450	0.356	0.432
256	32	8	0.184	0.270	0.167	0.241	0.183	0.261	0.359	0.452	0.335	0.464
512	128	4	0.198	0.272	0.216	0.296	0.350	0.412	0.592	0.638	0.622	0.662
512	64	8	0.092	0.146	0.134	0.188	0.104	0.188	0.372	0.490	0.400	0.518
1024	256	4	0.430	0.504	0.402	0.504	0.648	0.716	0.792	0.808	0.776	0.808
1024	128	8	0.160	0.226	0.136	0.200	0.230	0.290	0.506	0.608	0.500	0.586
2048	512	4	0.746	0.790	0.804	0.844	0.868	0.880	0.876	0.892	0.912	0.932
2048	256	8	0.434	0.492	0.454	0.510	0.534	0.578	0.670	0.754	0.678	0.744
4096	1024	4	0.932	0.940	0.930	0.940	0.943	0.953	0.982	0.985	0.992	0.992
4096	512	8	0.914	0.922	0.910	0.918	0.910	0.925	0.967	0.978	0.895	0.923

TABLE 8

Rejection frequencies of the test (4.19) *under the alternative (5.8) for different choices of the long-memory function d defined in (5.10)-(5.12). The innovations are standard normal distributed.*

	model (5.8) with d_1				model (5.8) with d_4			
	tvARMA(1,1)		tvFARIMA(1,d,1)		tvARMA(1,1)		tvFARIMA(1,d,1)	
h-step prediction	med	dev	med	dev	med	dev	med	dev
5	19.1	52.4	4.8	3.6	12.3	42.3	4.5	3.3
10	25.2	56.5	10.7	5.0	18.1	44.0	10.1	4.7
25	43.2	54.6	25.4	7.8	36.5	40.1	26.3	10.8

TABLE 9

Prediction error by a fit of tvARMA(1,1) and tvFARIMA(1,d,1) models (median and median absolute deviation obtained by 1000 *simulation runs).*

shows that the monotonicity of the power as a function of the integral $F = \int_0^1 d(u)\,du$ can also be observed in samples of realistic size.

Remark 5.1. It is well known that fitting tvFAR or tvFARIMA to tvAR or tvARMA models yields to confounded estimates of the AR/MA coefficients and the long-memory parameter. As a consequence the approximation of the nominal level becomes less accurate if the AR polynomial $|1+\sum_{j=1}^{k} a_j(u)e^{-i\lambda j}|^2$ has roots close to the unit disc. For example, motivated by a comment of a reviewer, we have conducted a further simulation study investigating a tvAR(1) model. These results are not depicted for the sake of brevity but they clearly show that the approximation of the nominal level of the new test is not accurate if the AR coefficients vary in the interval $(0.85, 1)$. In this case the level is overestimated, and the test (4.19) decides too often for a long-memory process.

5.2. Simulation of prediction error

In this subsection we investigate the question what one loses by fitting a short-range dependent non-stationary model to data that is truly non-stationary and long-range dependent. For this purpose we simulate from the tvFARIMA$(0, d, 1)$-process in (5.8) with long-memory functions d_1 and d_4 in (5.10) and (5.12), respectively. We separately fit a tvARMA(1,1) model and a tvFARIMA(1,d,1) model to the data and use the state space framework in Palma et al. (2013) in order to predict future values and then compare the prediction errors of these two fitted models. To be more precise we consider the sample size $T = 1024$

with block length $N = 256$ (resulting in $M = 4$ blocks) and use the local Whittle estimator from Section 4 to estimate on each block the locally varying AR and MA coefficients for the tvARMA(1,1) model and the AR, MA and long-memory parameters for the tvFARIMA$(1, d, 1)$ model. With these time-varying coefficients we use the Kalman filter equations in Palma et al. (2013) and calculate $5, 10$ and 25-step predictors with each of these two models. The prediction error is calculated by sum of squared residuals

$$\sum_{\ell=1}^{k} \left(X_{t,+\ell,T} - \widehat{X}_{t+\ell,T}\right)^2 \ , \ \ell = 5, 10, 25.$$

In Table 9 we display the median and median absolute deviation of the prediction errors obtained in 1000 simulation runs. We observe that the predictions, which take the long memory property into account are substantially more accurate.

5.3. Data examples

Testing: As an illustration we apply the new test to two different datasets, where in both examples the mean function has been estimated as described in Section 4. As pointed out in the previous section the quality of predictions can be improved, if long range dependence is present in non stationary data and considered in the predictions. For this reason the test proposed in this paper can be useful to obtain more accurate forcasts.

The first data set contains annual pinus longaeva tree ring width measurements at Mammoth Creek, Utah, from 0 A.D. to 1989 A.D. while the second data set contains 2048 squared log-returns of the IBM stock between July 15th 2005 and August 30th 2013 which was already discussed in the introduction. Both time series are depicted in Figure 5, and in the case of the tree ring data our test statistic $\sqrt{T}\hat{F}_T/\sqrt{\hat{W}_T}$ equals 17.8 for $M = 4$ and yields a p-value ≈ 0. This implies that the null hypothesis of a non-stationary short-memory model has to be rejected for this dataset, which coincides with the results of the tests in Baek and Pipiras (2012) and Yau and Davis (2012). Their test statistics have the values 3.49 and 9.37 and p-values of 0.00024 and 0 corresponding to the local Whittle and likelihood ratio approach, respectively. The CUSUM procedure of Berkes et al. (2006) yields a value of 0.906 for the test statistic and does not reject the null hypothesis at even 10% nominal level. This result is possibly due to the low power of this test as remarked in Section 5.1.

In the situation of the squared log-returns of the IBM stock, the assumption of Gaussianity is too restrictive and we therefore apply the more general test described in Remark 4.8. The values of the test statistic $\sqrt{T}\hat{F}_T/\sqrt{\hat{W}_{T,general}}$ are 5.67 and 9.48 for $M = 4$ and $M = 8$, respectively, yielding that the p-value is smaller than $2.87 \cdot 10^{-7}$ for both choices of the segmentation. This means that the assumption of no long-range dependence is clearly rejected. If we apply the likelihood ratio test of Yau and Davis (2012) to this dataset, we obtain a value for the statistic of 15.77 which is then compared with the quantiles

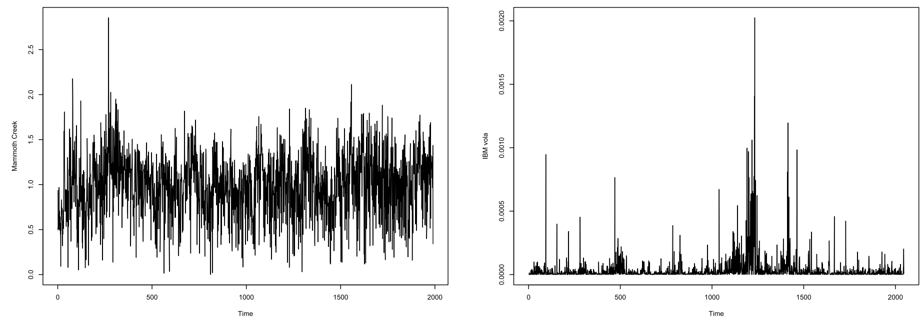

Fig 5. *Left panel: plot of the 1990 annual pinus longaeva tree ring width measurements at Mammoth Creek, Utah from 0 A.D. to 1989 A.D.; Right panel: plot of the squared log-returns of the IBM stock between July* 15*th* 2005 *and August* 30*th* 2013.

of the standard normal distribution. This yields also to a rejection of the null hypothesis. On the other hand, the CUSUM procedure of Berkes et al. (2006) only rejects the null hypothesis of no long-range dependence at a 10% but not at a 5% level. This observation is, however, not surprising given the low power of this test in the finite sample situations presented in the previous section. The test of Baek and Pipiras (2012) rejects the null hypothesis with a p-value $8.65 \cdot 10^{-12}$, yielding the same result as our approach and the one of Yau and Davis (2012).

Prediction: The result of the test (4.19) has important consequences for the subsequent data analysis as it advices the statistician to use short memory or long memory (non-stationary) models. In the final part of this section we demonstrate how the information of the test can be employed to obtain superior forecasting results in the two datasets analyzed in the previous paragraph. For this purpose, we divide both datasets into two parts. One part contains the first $3/4 \times T$ observations of the corresponding dataset while the second part contains the remaining $T/4$ data points. The new testing procedure (4.19) proposed in this paper is applied to the first part of the data, and - depending on the result of the test - forecasts are performed using either a tvFARIMA(1,d,1) or a ARMA(1,1) with the window of size $N = 256$ in the localized Whittle estimator [see also Section 5.2]. In order to compare the forecasting performance of the short- with the long-memory model, we define the prediction error on the second part of each dataset by

$$PE(h) = \sum_{t=3/4T+1}^{T} \sum_{\ell=1}^{h} \left(X_{t+\ell,T} - \widehat{X}_{t+\ell,T} \right)^2 , \; h = 5, 10, 25,$$

and denote with $PE_{short}(h)$ and $PE_{long}(h)$ the prediction error for the short- and long-memory approach respectively. The expression

$$R(h) = \frac{PE_{long}(h)}{PE_{short}(h)}$$

	$R(h)$		
dataset	$h = 5$	$h = 10$	$h = 25$
Mammoth Creek Data	0.21	0.24	0.26
IBM Data	0.09	0.15	0.33

TABLE 10

Comparison of prediction errors for the Mammoth Creek and IBM dataset with different values of h. A value smaller than one indicates a better performance of the long range dependent model.

then serves as a measure for the comparison. It is smaller than one if the long-range dependence approach yields superior predictions, while it is larger than one in the other case. As in the previous paragraph (where we applied the test to the total sample), an application of the test (4.19) to the first $3/4 \times T$ observations of the Mammoth Creek and the IBM dataset yields p-values much smaller than one percent in both cases. Consequently one would perform data analysis on the basis of a non-stationary long range dependent model. The advantages of this approach are clearly visible in Table 10 where we depict the ratio of the prediction error from a short and long range dependent model. We observe that the long-range dependence approach, in fact, yields substantially smaller prediction errors. In all cases the prediction error from the long-range dependent model is less than one third of the corresponding error from the short-memory model (for both datasets and all considered values of h). This demonstrates that the difference in forecasting performance is huge and highlights the importance of powerful tests to discriminate between long- and short-range dependence.

6. Conclusions

In this paper we have developed a test for weak against strong (long-range) dependence for non-stationarity time series. Our approach is based on an average of unconstrained Whittle-likelihood estimates of the (nonnegative) local long-range dependence parameter from a sequence of approximating time varying FARIMA models [see equation (4.12) for its definition]. It is demonstrated that a standardized version of this average is asymptotically normal distributed, which provides a very simple asymptotic level α and consistent test for discriminating between short and long range dependence of a non-stationary time series.

As an alternative to the statistic $\hat{F}_T$ in (4.12) one could form an average of constrained Whittle-likelihood estimates, say $\hat{d}_{N,c}(u_i)$. Constrained parameter estimation has found considerable attention in the literature [see for example Chernoff (1954) or Andrews (1999) among many others], but - to our best knowledge - it has not been considered so far in locally stationary processes. The "classical" results indicate that for a fixed value u the asymptotic distribution of $\hat{d}_{N,c}(u)$ is given by a function of a multivariate normal distribution (in the simplest case a half normal type distribution). However, we expect that - due to averaging - the (standardized) statistic $\hat{F}_{T,c} = \frac{1}{M}\sum_{i=1}^{M} \hat{d}_{N,c}(u_i)$ is still asymptotically normal distributed. An interesting direction for future research is the development of an asymptotic theory for constrained estimators in locally stationary (long memory) processes and to use it for a rigorous investigation of the asymptotic properties of the statistic $\hat{F}_{T,c}$. Moreover, the results of Section 3 for the nonparametric regression model with independent errors indicate some advantages of unconstrained over unconstrained averages, and it will be of interest to investigate if the superiority of $\hat{F}_T$ over $\hat{F}_{T,c}$ can also be observed for the testing problem considered in this paper.

It is also notable that this paper has its focus on discriminating between short and standard long-range dependence, which corresponds to a pole of the local spectral density at frequency 0. However, it was pointed out by several authors [see for example Arteche and Robinson (2000); Hidalgo and Soulier (2004); Reisen et al. (2006) among others] that - due to strong cyclic components - strong dependency can also occur as a pole in the spectral density at any other frequency (reflecting strong seasonal long range dependence). In this case the analogue of the model (2.6) is given by

$$f(u,\lambda) = |1-e^{i(\lambda-\lambda_0)}|^{-d_0(u)}|1-e^{i(\lambda+\lambda_0)}|^{-d_0(u)}g(u,\lambda), \tag{6.1}$$

where λ_0 denotes the unknown pole [see Hidalgo and Soulier (2004)], and a further interesting direction of future research is the construction of tests for the hypothesis (2.12) in the more general model (6.1).

We finally note that several authors have analyzed financial data under linearity assumptions as made in equation (2.1) [see Mikosch and Starica (2004), Perron and Qu (2010) or Chen et al. (2010) among others]. On the other hand it is also argued in the literature that this assumption might not be reasonable in some cases. Long range dependent processes have mainly been investigated in models with linear representations. A nonlinear (nonparametric) extension does not seem to be obvious as indicated by the results of Grublyte and Surgailis (2014), who proposed a linear representation with random coefficients. Therefore, an interesting problem for future research is to investigate if the methodology suggested in this paper is also valid for processes with nonlinear representations.

7. Appendix: Proofs

7.1. *Preliminary results*

We begin stating two results, which will be the main tools in the asymptotic analysis of the proposed estimators and the test statistic. For this purpose, we let $\phi_T : [0,1]\times[-\pi,\pi]\to\mathbb{R}$ denote a function which (might) depend on the the sample size T and define

$$\begin{aligned} G_T(\phi_T) &= \frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi} f(u_j,\lambda)\phi_T(u_j,\lambda)\,d\lambda, \\ \hat{G}_T(\phi_T) &= \frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi} I_N^{\mu}(u_j,\lambda)\phi_T(u_j,\lambda)\,d\lambda, \end{aligned}$$

where I_N^μ is the analogue of the local periodogram (4.4) where the estimator $\hat{\mu}$ has been replaced by the "true" mean function μ.

Theorem 7.1. *a) Let Assumption 2.1 be fulfilled and assume that $\phi_T(u,\lambda) : [0,1] \times [-\pi,\pi] \to \mathbb{R}$ is symmetric in λ, twice continuously differentiable with uniformly bounded partial derivatives such that for all $u \in [0,1]$, $\lambda \in [-\pi,\pi]$, $T \in \mathbb{N}$*

$$\begin{aligned} \phi_T(u,\lambda) &\leq Cg(k)|\lambda|^{2d_0(u)-\epsilon}, && (7.1)\\ \frac{\partial}{\partial\lambda}\phi_T(u,\lambda) &\leq Cg(k)|\lambda|^{2d_0(u)-1-\epsilon}, && (7.2)\\ \frac{\partial^2}{\partial\lambda^2}\phi_T(u,\lambda) &\leq Cg(k)|\lambda|^{2d_0(u)-2-\epsilon}, && (7.3)\end{aligned}$$

where $C>0, 0<\epsilon<1/2-D$ are constants and $g:\mathbb{N}\to(0,\infty)$ is a given function. Then we have

$$\begin{aligned} \mathbb{E}[\hat{G}_T(\phi_T)] &= G_T(\phi_T) + O\Big(\frac{g(k)}{N^{1-\epsilon}}\Big) + O\Big(\frac{g(k)N^2}{T^2}\Big), && (7.4)\\ \mathrm{Var}[\hat{G}_T(\phi_T)] &= V_T + O\Big(\frac{1}{T}\frac{g^2(k)}{N^{1-2D-2\epsilon}}\Big) + O\Big(\frac{g^2(k)N^2}{T^3}\Big) && (7.5)\end{aligned}$$

where

$$V_T = \frac{1}{T}\frac{4\pi}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi} f^2(u_j,\lambda)\phi_T^2(u_j,\lambda)\,d\lambda.$$

b) Suppose the assumptions of part a) hold with $D=0$, $\epsilon<1/6$ and additionally $\liminf_{T\to\infty} T\cdot V_T \geq c$,

$$\begin{aligned} &N\to\infty,\ g(k)\sqrt{T}/N^{1-\epsilon}\to 0,\ g(k)\log(T)/T^{1/6-\epsilon}\to 0,\\ &\textit{and}\ \ g(k)N^2/T^{\frac{3}{2}}\to 0.\end{aligned}$$

Then we have

$$\sqrt{T}\big(\hat{G}_T(\phi_T)-G_T(\phi_T)\big)/\sqrt{V_T} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Proof: In order to prove part a) Theorem 7.1 we define $\tilde{t}_j := t_j - N/2+1$, $\tilde{\psi}_l(u_{j,p}) := \psi_l(\frac{\tilde{t}_j+p}{T})$, $Z_{a,b} := Z_{a-N/2+1+b}$ and obtain

$$\begin{aligned} \mathbb{E}[\hat{G}_T(\phi_T)] &= \frac{1}{2\pi N}\frac{1}{M}\sum_{j=1}^{M}\sum_{p,q=0}^{N-1}\sum_{l,m=0}^{\infty}\psi_{\tilde{t}_j+p,T,l}\psi_{\tilde{t}_j+q,T,m}\\ &\qquad\times\int_{-\pi}^{\pi}\phi_T(u_j,\lambda)e^{-i(p-q)\lambda}\,d\lambda\,\mathbb{E}(Z_{t_j,p-l}Z_{t_j,q-m})\\ &= E_{N,T}+A_{N,T}+B_{N,T}\end{aligned}$$

where

$$
\begin{aligned}
E_{N,T} &:= \frac{1}{2\pi N}\frac{1}{M}\sum_{j=1}^{M}\sum_{p,q=0}^{N-1}\sum_{l,m=0}^{\infty}\psi_l(u_j)\psi_m(u_j) \\
&\qquad \times\int_{-\pi}^{\pi}\phi_T(u_j,\lambda)e^{-i(p-q)\lambda}\,d\lambda\mathbb{E}(Z_{t_j,p-l}Z_{t_j,q-m}),\\
A_{N,T} &:= \frac{1}{2\pi N}\frac{1}{M}\sum_{j=1}^{M}\sum_{p,q=0}^{N-1}\sum_{l,m=0}^{\infty}\int_{-\pi}^{\pi}\phi_T(u_j,\lambda)e^{-i(p-q)\lambda}\,d\lambda\mathbb{E}(Z_{t_j,p-l}Z_{t_j,q-m})\\
&\qquad \big\{\psi_l(u_j)\big(\tilde\psi_m(u_{j,q})-\psi_m(u_j)\big)+\big(\tilde\psi_l(u_{j,p})-\psi_l(u_j)\big)\tilde\psi_m(u_{j,q})\big\},\\
B_{N,T} &:= \frac{1}{2\pi N}\frac{1}{M}\sum_{j=1}^{M}\sum_{p,q=0}^{N-1}\sum_{l,m=0}^{\infty}\int_{-\pi}^{\pi}\phi_T(u_j,\lambda)e^{-i(p-q)\lambda}\,d\lambda\mathbb{E}(Z_{t_j,p-l}Z_{t_j,q-m})\\
&\qquad \big\{\big(\psi_{\tilde t_j+p,T,l}-\tilde\psi_l(u_{j,p})\big)\psi_{\tilde t_j+q,T,m}+\tilde\psi_l(u_{j,p})\big(\psi_{\tilde t_j+q,T,m}-\tilde\psi_m(u_{j,q})\big)\big\}.
\end{aligned}
$$

Note that $B_{N,T}$ and $A_{N,T}$ compromise the error arising in the approximation of $\psi_{\tilde t_j+p,T,l}$ by $\psi_l(\frac{\tilde t_j+p}{T})$ and $\tilde\psi_m(u_{j,q})$ by $\psi_m(u_j)$, respectively. In order to establish the claim (7.4), we prove the following statements:

$$
\begin{aligned}
E_{N,T} &= \frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}f(u_j,\lambda)\phi_T(u_j,\lambda)\,d\lambda+O\Big(\frac{g(k)}{N^{1-\epsilon}}\Big) &(7.6)\\
A_{N,T} &= O\Big(\frac{g(k)\log(N)}{N^{1-\epsilon}M}\Big)+O\Big(\frac{g(k)N^2}{T^2}\Big) &(7.7)\\
B_{N,T} &= O\Big(\frac{g(k)\log(T)}{T}\Big). &(7.8)
\end{aligned}
$$

Proof of (7.6): Due to the independence of the random variables Z_t, we only need to consider terms fulfilling $p=q+l-m$ (this means $0\le p=q+l-m\le N-1$ because of $p\in\{0,1,2\dots,N-1\}$) which in turn implies $|l-m|\le N-1$. Therefore

$$
\begin{aligned}
E_{N,T} &= \frac{1}{2\pi N}\frac{1}{M}\sum_{j=1}^{M}\sum_{\substack{l,m=0\\|l-m|\le N-1}}^{\infty}\sum_{\substack{q=0\\0\le q+l-m\le N-1}}^{N-1}\psi_l(u_j)\psi_m(u_j)\\
&\qquad\times\int_{-\pi}^{\pi}\phi_T(u_j,\lambda)e^{-i(l-m)\lambda}\,d\lambda\\
&= \frac{1}{2\pi N}\frac{1}{M}\sum_{j=1}^{M}\sum_{\substack{l,m=0\\|l-m|\le N-1}}^{\infty}\psi_l(u_j)\psi_m(u_j)\\
&\qquad\times\int_{-\pi}^{\pi}\phi_T(u_j,\lambda)e^{-i(l-m)\lambda}\,d\lambda(N-|l-m|)\\
&= \frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\phi_T(u_j,\lambda)f(u_j,\lambda)\,d\lambda+E^1_{N,T}+E^2_{N,T},
\end{aligned}
$$

where

$$E^1_{N,T} = -\frac{1}{2\pi}\frac{1}{M}\sum_{j=1}^{M}\sum_{\substack{l,m=0\\ N\le|l-m|}}^{\infty}\psi_l(u_j)\psi_m(u_j)\int_{-\pi}^{\pi}\phi_T(u_j,\lambda)e^{-i(l-m)\lambda}\,d\lambda,$$

$$E^2_{N,T} = \frac{-1}{2\pi N}\frac{1}{M}\sum_{j=1}^{M}\sum_{\substack{l,m=0\\ |l-m|\le N-1}}^{\infty}\psi_l(u_j)\psi_m(u_j)\int_{-\pi}^{\pi}\phi_T(u_j,\lambda)e^{-i(l-m)\lambda}\,d\lambda|l-m|.$$

Using (2.4), (7.1) and Lemma 8.2 in the supplement, we obtain

$$|E^1_{N,T}| \le C\frac{g(k)}{M}\sum_{j=1}^{M}\sum_{\substack{l,m=1\\ N\le|l-m|}}^{\infty}\frac{1}{l^{1-d_0(u_j)}}\frac{1}{m^{1-d_0(u_j)}}\frac{1}{|l-m|^{1+2d_0(u_j)-\epsilon}}(1+o(1)),$$

where we used the fact that terms corresponding to $l=0$ or $m=0$ are of smaller or the same order (we will use this property frequently from now on without further mentioning it). We set $h := l-m$ and obtain from Lemma 8.1a) in the online supplement that

$$\frac{g(k)}{M}\sum_{j=1}^{M}\sum_{\substack{h\in\mathbb{Z}\\ N\le|h|}}\sum_{\substack{m=1\\ h+m\ge1}}^{\infty}\frac{1}{(h+m)^{1-d_0(u_j)}}\frac{1}{m^{1-d_0(u_j)}}\frac{1}{|h|^{1+2d_0(u_j)-\epsilon}}$$
$$\le Cg(k)\sum_{\substack{h\in\mathbb{Z}\\ N\le|h|}}\frac{1}{|h|^{2-\epsilon}} = O\Big(\frac{g(k)}{N^{1-\epsilon}}\Big).$$

By proceeding analogously we obtain that $E^2_{N,T} = O(g(k)N^{-1+\epsilon})$ which proves the assertion in (7.6).

Proof of (7.7): Without loss of generality we only consider the first summand

$$A_{N,T}(1) = \frac{1}{2\pi N}\frac{1}{M}\sum_{j=1}^{M}\sum_{p,q=0}^{N-1}\sum_{l,m=0}^{\infty}\psi_l(u_j)\big(\tilde{\psi}_m(u_{j,q})-\psi_m(u_j)\big)$$
$$\times\int_{-\pi}^{\pi}\phi_T(u_j,\lambda)e^{-i(p-q)\lambda}\,d\lambda\mathbb{E}(Z_{t_j,p-l}Z_{t_j,q-m})$$

in $A_{N,T}$ (the second term is treated exactly in the same way). A Taylor expansion and similar arguments as in the proof of (7.6) yield

$$A_{N,T}(1) = A^1_{N,T} + A^2_{N,T}$$

where

$$A^1_{N,T} = \frac{1}{2\pi N}\frac{1}{M}\sum_{j=1}^{M}\sum_{\substack{l,m=0\\ |l-m|\le N-1}}^{\infty}\sum_{\substack{q=0\\ 0\le q+l-m\le N-1}}^{N-1}\psi_l(u_j)\psi'_m(u_j)$$

$$\times \Big(\frac{-N/2+1+q}{T}\Big) \int_{-\pi}^{\pi} \phi_T(u_j,\lambda) e^{-i(l-m)\lambda}\, d\lambda,$$

$$A^2_{N,T} = \frac{1}{2\pi N}\frac{1}{M}\sum_{j=1}^{M} \sum_{\substack{l,m=0\\ |l-m|\le N-1}}^{\infty} \sum_{\substack{q=0\\ 0\le q+l-m\le N-1}}^{N-1} \psi_l(u_j)\psi_m^{''}(\eta_{m,j,q}) \times \Big(\frac{-N/2+1+q}{T}\Big)^2 \int_{-\pi}^{\pi} \phi_T(u_j,\lambda) e^{-i(l-m)\lambda}\, d\lambda$$

and $\eta_{m,j,q} \in (u_j - N/(2T), u_j + N/(2T))$. Using (2.4), (2.8), (7.1), Lemma 8.2 it follows

$$\begin{aligned}
|A^1_{N,T}| &\le C\frac{g(k)}{N}\frac{1}{M}\sum_{j=1}^{M} \sum_{\substack{l,m=1\\ 1\le|l-m|\le N-1}}^{\infty} \frac{1}{l^{1-d_0(u_j)}}\frac{\log(m)}{m^{1-d_0(u_j)}}\frac{1}{|l-m|^{1+2d_0(u_j)-\epsilon}} \\
&\qquad \times \Big| \sum_{\substack{q=0\\ 0\le q+l-m\le N-1}}^{N-1} \Big(\frac{-N/2+1+q}{T}\Big)\Big| \\
&\le C\frac{g(k)}{T}\frac{1}{M}\sum_{j=1}^{M} \sum_{\substack{l,m=1\\ 1\le|l-m|\le N-1}}^{\infty} \frac{1}{l^{1-d_0(u_j)}}\frac{\log(m)}{m^{1-d_0(u_j)}}\frac{1}{|l-m|^{2d_0(u_j)-\epsilon}} \\
&= C\frac{g(k)}{T}\frac{1}{M}\sum_{j=1}^{M} \sum_{\substack{s\in\mathbb{Z}\\ 1\le|s|\le N-1}} \sum_{\substack{l=1\\ 1\le l-s}}^{\infty} \frac{1}{l^{1-d_0(u_j)}}\frac{\log(l-s)}{(l-s)^{1-d_0(u_j)}}\frac{1}{|s|^{2d_0(u_j)-\epsilon}} \\
&\le C\frac{g(k)\log(N)}{T}\frac{1}{M}\sum_{j=1}^{M} \sum_{\substack{s\in\mathbb{Z}\\ 1\le|s|\le N-1}} \frac{1}{|s|^{1-\epsilon}} = O\Big(\frac{g(k)\log(N)}{N^{1-\epsilon}M}\Big)
\end{aligned}$$

where we used Lemma 8.1(c) in the online supplement for the last step. Finally, (2.4), (2.8), (7.1), Lemma 8.2 in the online supplement and the same arguments as above, show that the term $A^2_{N,T}$ is of order $O(g(k)N^2T^{-2})$.

Proof of (7.8): By employing (2.3) and the same arguments as above it can be shown that $B_{N,T}$ is of order $O(\frac{g(k)\log(T)}{T})$.

In the next step we prove the asymptotic representation for the variance in (7.5). We obtain

$$\begin{aligned}
\mathrm{Var}(\hat{G}_T(\phi_T)) &= \frac{1}{(2\pi N)^2}\frac{1}{M^2}\sum_{j_1,j_2=1}^{M}\sum_{p,q,r,s=0}^{N-1}\sum_{l,m,n,o=0}^{\infty} \psi_l(u_{j_1})\psi_m(u_{j_1})\psi_n(u_{j_2})\psi_o(u_{j_2}) \\
&\quad \times \mathrm{cum}(Z_{t_{j_1},p-l}Z_{t_{j_1},q-m}, Z_{t_{j_2},r-n}Z_{t_{j_2},s-o}) \\
&\quad \times \int_{-\pi}^{\pi} \phi_T(u_{j_1},\lambda_1)e^{-i(p-q)\lambda_1}\, d\lambda_1 \int_{-\pi}^{\pi} \phi_T(u_{j_2},\lambda_2)e^{-i(r-s)\lambda_2}\, d\lambda_2
\end{aligned}$$

$$+ O\Big(\frac{g^2(k)\log(N)}{TN^{1-\epsilon}M}\Big) + O\Big(\frac{g^2(k)N^2}{T^3}\Big),$$

where we used assumption (2.3) and similar arguments as given in the proof of (7.4). Because of the Gaussianity of the innovations we obtain

$$\begin{aligned}\mathrm{cum}(Z_{t_{j_1},p-l}Z_{t_{j_1},q-m}, Z_{t_{j_2},r-n}Z_{t_{j_2},s-o}) = \\ &\mathbb{E}(Z_{t_{j_1},p-l}Z_{t_{j_2},r-n})\mathbb{E}(Z_{t_{j_1},q-m}Z_{t_{j_2},s-o}) \\ &+ \mathbb{E}(Z_{t_{j_1},p-l}Z_{t_{j_2},s-o})\mathbb{E}(Z_{t_{j_1},q-m}Z_{t_{j_2},r-n}).\end{aligned}$$

This implies that the calculation of the (dominating part of the) variance splits into two sums, say $V^1_{N,T}$ and $V^2_{N,T}$. In the following discussion we will show that both terms converge to the same limit, that is

$$V^i_{N,T} = \frac{1}{T}\frac{2\pi}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi} f^2(u_j,\lambda)\phi_T^2(u_j,\lambda)\,d\lambda + O\Big(\frac{1}{T}\frac{g^2(k)}{N^{1-2D-2\epsilon}}\Big); \quad i = 1,2$$

For the sake of brevity we restrict ourselves to the case $i = 1$. Because of the independence of the innovations Z_t, we obtain that the conditions $p = r + l - n + (j_2 - j_1)N$ and $s = q + o - m + (j_1 - j_2)N$ must hold, which, because of $p, s \in \{0, ..., N-1\}$, directly implies $|l - n + (j_2 - j_1)N| \le N - 1$ and $|o - m + (j_1 - j_2)N| \le N - 1$. Thus, the term $V^1_{N,T}$ can be written as

$$\begin{aligned}\frac{1}{(2\pi N)^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{q,r=0}^{N-1}\sum_{l,m,n,o=0}^{\infty}\sum_{\substack{j_2=1 \\ 0\le r+l-n+(j_2-j_1)N\le N-1 \\ 0\le q+o-m+(j_1-j_2)N\le N-1 \\ |l-n+(j_2-j_1)N|\le N-1 \\ |o-m+(j_1-j_2)N|\le N-1}}^{M} \psi_l(u_{j_1})\psi_m(u_{j_1})\psi_n(u_{j_2})\psi_o(u_{j_2}) \\ \times \int_{-\pi}^{\pi}\phi_T(u_{j_1},\lambda_1)e^{-i(r-q+l-n+(j_2-j_1)N)\lambda_1}\,d\lambda_1 \\ \times \int_{-\pi}^{\pi}\phi_T(u_{j_2},\lambda_2)e^{-i(r-q+m-o+(j_2-j_1)N)\lambda_2}\,d\lambda_2.\end{aligned}$$

Since $q \in \{0, 1, 2\ldots, N-1\}$, we get from the condition $0 \le q + o - m + (j_1 - j_2)N \le N - 1$ that, if q, o, m, j_1 are fixed, there are at most two possible values for j_2 such that the corresponding term does not vanish. It follows from Lemma 8.3 (i)–(iii) in the online supplement that there appears an error of order $O(\frac{1}{T}\frac{g^2(k)}{N^{1-2D-2\epsilon}})$ if we drop the condition $0 \le r + l - n + (j_2 - j_1)N \le N - 1$ and assume that the variable r runs from $-(N-1)$ to -1. Therefore, up to an error of order $O(\frac{1}{T}\frac{g^2(k)}{N^{1-2D-2\epsilon}})$, the term $V^1_{N,T}$ is equal to

$$D_{1,T} + D_{2,T},$$

where

$$D_{1,T} = \frac{1}{(2\pi N)^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{q=0}^{N-1}\sum_{r=-(N-1)}^{N-1}\sum_{l,m,n,o=0}^{\infty}$$

$$\sum_{\substack{j_2=1\\ 0\le q+o-m+(j_1-j_2)N\le N-1\\ |l-n+(j_2-j_1)N|\le N-1\\ |o-m+(j_1-j_2)N|\le N-1}}^{M} \psi_l(u_{j_1})\psi_m(u_{j_1})\psi_n(u_{j_2})\psi_o(u_{j_2})$$

$$\times \int_{-\pi}^{\pi} \phi_T(u_{j_1},\lambda_1)\phi_T(u_{j_2},\lambda_1)e^{-i(r-q+l-n+(j_2-j_1)N)\lambda_1}\, d\lambda_1$$

$$\times \int_{-\pi}^{\pi} e^{-i(r-q+m-o+(j_2-j_1)N)\lambda_2}\, d\lambda_2$$

$$D_{2,T} = \frac{1}{(2\pi N)^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{q=0}^{N-1}\sum_{r=-(N-1)}^{N-1}\sum_{l,m,n,o=0}^{\infty}$$

$$\times \sum_{\substack{j_2=1\\ 0\le q+o-m+(j_1-j_2)N\le N-1\\ |l-n+(j_2-j_1)N|\le N-1\\ |o-m+(j_1-j_2)N|\le N-1}}^{M} \psi_l(u_{j_1})\psi_m(u_{j_1})\psi_n(u_{j_2})\psi_o(u_{j_2})$$

$$\times \int_{-\pi}^{\pi} \phi_T(u_{j_1},\lambda_1)e^{-i(r-q+l-n+(j_2-j_1)N)\lambda_1}$$

$$\times \int_{-\pi}^{\pi} \Big[\phi_T(u_{j_2},\lambda_2)-\phi_T(u_{j_2},\lambda_1)\Big]e^{-i(r-q+m-o+(j_2-j_1)N)\lambda_2}\, d\lambda_2\, d\lambda_1.$$

We show

$$D_{1,T} = \frac{2\pi}{N}\frac{1}{M^2}\sum_{j_1=1}^{M}\int_{-\pi}^{\pi} f^2(u_{j_1},\lambda_1)\phi_T^2(u_{j_1},\lambda_1)\, d\lambda_1 + O\Big(\frac{1}{T}\frac{g^2(k)}{N^{1-2D-2\epsilon}}\Big) \quad (7.9)$$

$$D_{2,T} = O\Big(\frac{1}{T}\frac{g^2(k)}{N^{1-2D-2\epsilon}}\Big),$$

which then concludes the proof of (7.5). For this purpose we begin with an investigation of the term $D_{1,T}$ for which the terms in the sum vanish if $r-q+m-o+(j_2-j_1)N\neq 0$. Moreover, the following facts are correct:

I. The variable r runs from 0 to $N-1$ since $r-q+m-o+(j_2-j_1)N=0$ and $0\le q+o-m+(j_1-j_2)N\le N-1$.
II. We can drop the condition $|l-n+(j_2-j_1)N|\le N-1$ by making an error of order $O(g^2(k)T^{-1}N^{-1+2D+2\epsilon})$ [this follows from Lemma 8.3(iv) in the online supplement].
III. There appears an error of order $O(g^2(k)T^{-1}N^{-1+2D+2\epsilon})$ if we omit the sum with $j_1\neq j_2$ [we prove this in Lemma 8.3(v) in the online supplement].
IV. We can afterwards omit the condition $0\le q+o-m\le N-1$ since it is $0\le r\le N-1$ and $r-q+m-o=0$ [note that, because of III., we assume $j_1=j_2$ from now on].
V. We can then drop the condition $|o-m|\le N-1$ since $r-q+m-o=0$ and $|r-q|\le N-1$.

Thus, using the representation of $f(u_{j_1},\lambda)$ in (2.5), the term $D_{1,T}$ can be written as (up to an error of order $O(g^2(k)T^{-1}N^{-1+2D+2\epsilon})$)

$$\begin{aligned}
& \frac{1}{N^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{q,r=0}^{N-1}\int_{-\pi}^{\pi} f(u_{j_1},\lambda_1)\phi_T^2(u_{j_1},\lambda_1)e^{-i(r-q)\lambda_1}\,d\lambda_1 \\
& \qquad\qquad \times \int_{-\pi}^{\pi} f(u_{j_1},\lambda_2)e^{-i(r-q)\lambda_2}\,d\lambda_2 \\
= & \frac{1}{N^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{s=-(N-1)}^{N-1}\int_{-\pi}^{\pi} f(u_{j_1},\lambda_1)\phi_T^2(u_{j_1},\lambda_1)e^{-is\lambda_1}\,d\lambda_1 \\
& \qquad\qquad \times \int_{-\pi}^{\pi} f(u_{j_1},\lambda_2)e^{-is\lambda_2}\,d\lambda_2(N-|s|) \\
= & D_{1,T}^{(1)} + D_{1,T}^{(2)} + D_{1,T}^{(3)},
\end{aligned}$$

where

$$\begin{aligned}
D_{1,T}^{(1)} &= \frac{1}{N}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{s=-\infty}^{\infty}\int_{-\pi}^{\pi} f(u_{j_1},\lambda_1)\phi_T^2(u_{j_1},\lambda_1)e^{-is\lambda_1}\,d\lambda_1 \\
& \qquad\qquad \times \int_{-\pi}^{\pi} f(u_{j_1},\lambda_2)e^{-is\lambda_2}\,d\lambda_2 \\
D_{1,T}^{(2)} &= -\frac{1}{N}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{\substack{s\in\mathbb{Z}\\|s|\geq N}}\int_{-\pi}^{\pi} f(u_{j_1},\lambda_1)\phi_T^2(u_{j_1},\lambda_1)e^{-is\lambda_1}\,d\lambda_1 \\
& \qquad\qquad \times \int_{-\pi}^{\pi} f(u_{j_1},\lambda_2)e^{-is\lambda_2}\,d\lambda_2 \\
D_{1,T}^{(3)} &= -\frac{1}{N^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{s=-(N-1)}^{N-1}|s|\int_{-\pi}^{\pi} f(u_{j_1},\lambda_1)\phi_T^2(u_{j_1},\lambda_1)e^{-is\lambda_1}\,d\lambda_1 \\
& \qquad\qquad\qquad \times \int_{-\pi}^{\pi} f(u_{j_1},\lambda_2)e^{-is\lambda_2}\,d\lambda_2
\end{aligned}$$

With Parseval's identity, we get

$$D_{1,T}^{(1)} = \frac{2\pi}{N}\frac{1}{M^2}\sum_{j_1=1}^{M}\int_{-\pi}^{\pi} f^2(u_{j_1},\lambda_2)\phi_T^2(u_{j_1},\lambda_2)\,d\lambda_2,$$

while Lemma 8.2 in the online supplement yields (up to a constant) the inequalities

$$D_{1,T}^{(2)} \leq \frac{g^2(k)}{N}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{\substack{s\in\mathbb{Z}\\|s|\geq N}}\frac{1}{|s|^{2-2\epsilon}} \leq \frac{g^2(k)}{N^{2-2\epsilon}}\frac{1}{M},$$

$$D_{1,T}^{(3)} \leq \frac{g^2(k)}{N^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{\substack{s\in\mathbb{Z}\\1\leq|s|\leq N-1}}^{N-1}\frac{1}{|s|^{1-2\epsilon}} \leq \frac{g^2(k)}{N^{2-2\epsilon}M},$$

which proves (7.9). We now consider the term

$$D_{2,T} = D_{2,T}^{(1)} + D_{2,T}^{(2)},$$

where

$$\begin{aligned} D_{2,T}^{(1)} &= \frac{1}{(2\pi N)^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{q=0}^{N-1}\sum_{r=-\infty}^{\infty}\sum_{l,m,n,o=0}^{\infty} \\ &\qquad \sum_{\substack{j_2=1\\0\leq q+o-m+(j_1-j_2)N\leq N-1\\|l-n+(j_2-j_1)N|\leq N-1\\|o-m+(j_1-j_2)N|\leq N-1}}^{\infty} \psi_l(u_{j_1})\psi_m(u_{j_1})\psi_n(u_{j_2})\psi_o(u_{j_2}) \\ &\qquad \times\int_{-\pi}^{\pi}\phi_T(u_{j_1},\lambda_1)e^{-i(r-q+l-n+(j_2-j_1)N)\lambda_1} \\ &\qquad \times\int_{-\pi}^{\pi}\big[\phi_T(u_{j_2},\lambda_2)-\phi_T(u_{j_2},\lambda_1)\big]e^{-i(r-q+m-o+(j_2-j_1)N)\lambda_2}\,d\lambda_2\,d\lambda_1 \\ D_{2,T}^{(2)} &= -\frac{1}{(2\pi N)^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{q=0}^{N-1}\sum_{\substack{r\in\mathbb{Z}\\|r|\geq N}}\sum_{l,m,n,o=0}^{\infty} \\ &\qquad \sum_{\substack{j_2=1\\0\leq q+o-m+(j_1-j_2)N\leq N-1\\|l-n+(j_2-j_1)N|\leq N-1\\|o-m+(j_1-j_2)N|\leq N-1}}^{\infty} \psi_l(u_{j_1})\psi_m(u_{j_1})\psi_n(u_{j_2})\psi_o(u_{j_2}) \\ &\qquad \times\int_{-\pi}^{\pi}\phi_T(u_{j_1},\lambda_1)e^{-i(r-q+l-n+(j_2-j_1)N)\lambda_1} \\ &\qquad \times\int_{-\pi}^{\pi}\big[\phi_T(u_{j_2},\lambda_2)-\phi_T(u_{j_2},\lambda_1)\big]e^{-i(r-q+m-o+(j_2-j_1)N)\lambda_2}\,d\lambda_2\,d\lambda_1. \end{aligned}$$

Here $D_{2,T}^{(1)}$ corresponds to the sum over all r and vanishes by Parseval's identity. $D_{2,T}^{(2)}$ stands for the resulting error term which is of order $O(T^{-1}g^2(k)N^{-1+2D+2\epsilon})$ because of Lemma 8.3 (vi) in the online supplement.

Part b) follows with par a) if we show

$$\operatorname{cum}_l[\sqrt{T}\hat{G}_T(\phi)] = O\big(g(k)^l T^{l(\epsilon-1/2+2D)+(1-4D)}\log(T)^l\big) \qquad (7.10)$$

for $l \geq 3$ and $D < 1/4$. For a proof of this statement where we proceed (with a slight modification) analogously to the proof of Theorem 6.1 c) in Preuß and Vetter (2013). Note that these authors work with functions ϕ_T such that

$$\frac{1}{N}\sum_{k=1}^{N/2}\phi_T(u,\lambda_k)e^{ih\lambda_k} \quad = \quad O\Big(\frac{1}{|h \text{ modulo } N/2|}\Big) \tag{7.11}$$

while $\int_{-\pi}^{\pi}\phi_T(u,\lambda)e^{ih\lambda}d\lambda = O(h^{-1})$ for the integrated case. The authors then derive the exact same order as in (7.10) with the only difference that $\epsilon = 0$ and $g(k) \equiv 1$. In our situation, assumption (7.1) and Lemma 8.2 in the online supplement imply

$$\int_{-\pi}^{\pi}\phi_T(u,\lambda)e^{ih\lambda}\,d\lambda \quad = \quad O\Big(\frac{g(k)}{|h|^{1+2d_0(u)-\epsilon}}\Big) = O\Big(T^{\epsilon}\frac{g(k)}{|h|}\Big) \tag{7.12}$$

and we can therefore proceed completely analogously to the proof of Theorem 6.1 c) in Preuß and Vetter (2013) but using (7.12) instead of (7.11). The details are omitted for the sake of brevity. □

For the formulation of the next result we define the set

$$\begin{aligned}\mathcal{G}_T(s,\ell) \quad = \quad \big\{&\tilde{\phi}_T : [-\pi,\pi] \to \mathbb{R} \mid \tilde{\phi}_T \text{ is symmetric, there exists}\\ &\text{a polynomial } P_\ell \text{ of degree } \ell \text{ and a constant}\\ &d \in [-\gamma_k, 1/2) \text{ such that}\\ &\tilde{\phi}_T(\lambda) = \log^s(|1-e^{i\lambda}|)|1-e^{i\lambda}|^{2d}|P_\ell(e^{i\lambda})|^2\big\}\end{aligned}$$

and state the following result.

Theorem 7.2. *Suppose Assumption 2.1 and 4.2 are fulfilled, $N^{5/2}/T^2 \to 0$ and $0 < \epsilon < 1/4 - D/2$ is the constant of Assumption 4.2. Let Φ_T denote a class of functions $\phi_T : [0,1] \times [-\pi,\pi] \to \mathbb{R}$ consisting of elements, which are twice continuously differentiable with uniformly bounded partial derivates with respect to u, λ, T and satisfy (7.1)–(7.3) with $g(k) \equiv 1$, where the constant C does not depend on Φ_T, T. Furthermore, we assume that for all $u \in [0,1]$ the condition $\phi_T(u,\cdot) \in \mathcal{G}_T(s,qk)$ holds, where $q,s \in \mathbb{N}$ are fixed and $k = k(T)$ denotes a sequence satisfying $k^4\log^2(T)N^{-\epsilon/2} \to 0$. Then*

$$\sup_{u\in[0,1]}\sup_{\phi_T\in\Phi_T}\Big|\int_{-\pi}^{\pi}(I_N^{\mu}(u,\lambda) - f(\lfloor uT\rfloor/T,\lambda))\phi_T(\lfloor uT\rfloor/T,\lambda)d\lambda\Big| = o_P(N^{-1/2+\epsilon/2}).$$

Proof: We define Φ_T^* as the set of functions which we obtain by multiplying all elements $\phi_T \in \Phi_T$ with $1_{\{u=t/T\}}(u,\lambda)$, that is $\phi_T^*(u,\lambda) = \phi_T(t/T,\lambda)$ for some $t = 1,...,T$ and $\phi_T \in \Phi_T$, and consider

$$\hat{D}_{T,1}(\phi_T^*) := \sum_{t_1=1}^{T}\int_{-\pi}^{\pi}I_N^{\mu}(t_1/T,\lambda)\phi_T^*(t_1/T,\lambda)d\lambda, \quad \phi_T^* \in \Phi_T^*.$$

It follows from Theorem 2.1 in Newey (1991) that the assertion of Theorem 7.2 is a consequence of the statements:

(i) For every $\phi_T^* \in \Phi_T^*$ we have

$$\hat{G}_{T,1}(\phi_T^*) := N^{1/2-\epsilon/2}\Big(\hat{D}_{T,1}(\phi_T^*) - \int_{-\pi}^{\pi}f(t/T,\lambda)\phi_T(t/T,\lambda)d\lambda\Big) = o_p(1) \tag{7.13}$$

(ii) For every $\eta > 0$ we have

$$\lim_{T\to\infty} P\Big(\sup_{\phi^*_{T,1},\phi^*_{T,2}\in\Phi^*_T} |\hat{G}_{T,1}(\phi^*_{T,1}) - \hat{G}_{T,1}(\phi^*_{T,2})| > \eta\Big) = 0. \qquad (7.14)$$

In order to prove part (i) we use the same arguments as given in the proof of (7.4) and (7.5) and obtain

$$\begin{aligned}\mathbb{E}[\hat{D}_{T,1}(\phi^*_T))] &= \int_{-\pi}^{\pi} f(t/T,\lambda)\phi_T(t/T,\lambda)d\lambda + O\Big(\frac{1}{N^{1-\epsilon-2\gamma_K}}\Big) + O\Big(\frac{N^2}{T^2}\Big),\\ \mathrm{Var}[N^{1/2}\hat{D}_{T,1}(\phi^*_T)] &= \int_{-\pi}^{\pi} f^2(t/T,\lambda)\phi^2_T(t/T,\lambda)d\lambda + O\Big(\frac{1}{N^{1-2D-2\epsilon-4\gamma_k}}\Big)\\ &\quad + O\Big(\frac{N^2}{T^2}\Big),\end{aligned}$$

which yields (7.13) observing the growth conditions on N and T. For the proof of part (ii) we note that it follows by similar arguments as given in the proof of Theorem 6.1 d) of Preuß and Vetter (2013) that there exists a positive constant C such that the inequality

$$\mathbb{E}(|\hat{G}_{T,1}(\phi^*_{T,1}) - \hat{G}_{T,1}(\phi^*_{T,2})|^l) \leq (2l)!C^l\Delta^l_{T,\epsilon}(\phi^*_{T,1},\phi^*_{T,2})$$

holds for all even $l \in \mathbb{N}$ and all $\phi^*_{T,1},\phi^*_{T,2} \in \Phi^*_T$, where

$$\begin{aligned}\Delta_{T,\epsilon}(\phi^*_{T,1},\phi^*_{T,2}) \quad &= \quad 1_{\{t_1=t_2\}}N^{-\epsilon/2}\sqrt{\int_{-\pi}^{\pi}(\phi_{T,1,1}(t_1/T,\lambda) - \phi_{T,1,2}(t_1/T,\lambda))^2 d\lambda}\\ &\quad\quad +A1_{\{t_1\neq t_2\}}N^{-\epsilon/2}\end{aligned}$$

for a constant A which is sufficiently large such that

$$\sup_{\phi_{T,1,i}\in\Phi^*_T}\sqrt{\int_{-\pi}^{\pi}(\phi_{T,1,1}(t_1/T,\lambda) - \phi_{T,1,2}(t_1/T,\lambda))^2 d\lambda} \leq A.$$

By an application of Markov's inequality and a straightforward but cumbersome calculation [see the proof of Lemma 2.3 in Dahlhaus (1988) for more details] this yields

$$P(|\hat{G}_{T,1}(\phi^*_{T,1}) - \hat{G}_{T,1}(\phi^*_{T,2})| > \eta) \leq 96\exp(-\sqrt{\eta\Delta^{-1}_{T,\epsilon}(\phi^*_{T,1},\phi^*_{T,2})C^{-1}})$$

for all $\phi^*_{T,1},\phi^*_{T,2} \in \Phi^*_T$. The statement (7.14) then follows with the extension of the classical chaining argument as described in Dahlhaus (1988) if we show that the corresponding covering integral of Φ^*_T with respect to the semi-metric $\Delta_{T,\epsilon}$ is finite. More precisely, the covering number $N_T(u)$ of Φ^*_T with respect to $\Delta_{T,\epsilon}$ is equal to one for $u \geq AN^{-\epsilon/2}$ and bounded by $TC^{(qk)^2}u^{-qk}N^{-qk\epsilon/2}$ for some constant C for $u < AN^{-\epsilon/2}$ [see Chapter VII.2. of Pollard (1984) for a definition of covering numbers]. This implies that the covering integral $J_T(\delta) = \int_0^\delta[\log(48N_T(u)^2u^{-1}]^2du$ is up to a constant bounded by $k^4\log^2(T)N^{-\epsilon/2}$. The assertion follows by the assumptions on k and N. □

7.2. Proof of Theorem 4.3

Introducing the notation

$$\mathcal{L}^{\mu}_{N,k}(\theta_k, u) := \frac{1}{4\pi}\int_{-\pi}^{\pi}\Big(\log(f_{\theta_k}(\lambda)) + \frac{I^{\mu}_N(u,\lambda)}{f_{\theta_k}(\lambda)}\Big)\,d\lambda, \quad u \in [0,1]$$

we obtain with the same arguments as given in the proof of Theorem 3.6 in Dahlhaus (1997)

$$\max_{t=1,\dots,T}\big|\mathcal{L}^{\hat\mu}_{N,k}(\theta_k, t/T) - \mathcal{L}^{\mu}_{N,k}(\theta_k, t/T)\big| \le CN^{\epsilon}\max_{t=1,\dots,T}\big|\mu(t/T) - \hat\mu(t/T)\big|^2$$
$$+C\max_{t=1,\dots,T}\max_{q=0,\dots,N}\Big\{\big|\mu(t/T)-\hat\mu(t/T)\big|\Big|\int_{-\pi}^{\pi} d_N^{X-\mu}(t/T,\lambda) f^{-1}_{\theta_k}(\lambda)e^{iq\lambda}\,d\lambda\Big|\Big\}$$

for some constant $C \in \mathbb{R}$ and $d_N^{X-\mu}$ is defined by $\big|d_N^{X-\mu}(u,\lambda)\big|^2 := I^{\mu}_N(u,\lambda)$. By proceeding as in the proof of Theorem 7.2 one verifies

$$\max_{t=1,\dots,T}\max_{q=0,\dots,N}\sup_{\theta_k\in\Theta_{R,k}}\Big|\int_{-\pi}^{\pi} d_N^{X-\mu}(t/T,\lambda)f^{-1}_{\theta_k}(\lambda)e^{iq\lambda}\,d\lambda\Big| = O(N^{\epsilon}),$$

and (4.9) yields

$$\begin{aligned}
&\max_{t=1,\dots,T}\sup_{\theta_k\in\Theta_{R,k}}\big|\mathcal{L}^{\hat\mu}_{N,k}(\theta_k, t/T) - \mathcal{L}^{\mu}_{N,k}(\theta_k, t/T)\big| \\
&\quad = \max_{t=1,\dots,T}\big|\mu(t/T)-\hat\mu(t/T)\big|O_p(N^{\epsilon}) \\
&\quad = o_p(k^{-5/2}), \qquad (7.15)
\end{aligned}$$

and analogously we get

$$\begin{aligned}
&\max_{t=1,\dots,T}\sup_{\theta_k\in\Theta_{R,k}}\big\|\nabla\mathcal{L}^{\hat\mu}_{N,k}(\theta_k, t/T) - \nabla\mathcal{L}^{\mu}_{N,k}(\theta_k, t/T)\big\|_2 \qquad (7.16)\\
&\qquad = \max_{t=1,\dots,T}\big|\mu(t/T)-\hat\mu(t/T)\big|O_p(k^{1/2}N^{\epsilon}) = o_p(k^{-5/2}).
\end{aligned}$$

For each $u \in [0,1]$ let $\hat\theta_{N,k}(u)$ denote the Whittle-estimator defined in (4.2). Then Theorem 7.2 and similar arguments as in the proof of Theorem 3.2 in Dahlhaus (1997) yield

$$\sup_{u\in[0,1]}\big\|\hat\theta_{N,k}(u) - \theta_{0,k}(u)\big\|_2 \quad = \quad o_p(1). \qquad (7.17)$$

We will now derive a refinement of this statement. By an application of the mean value theorem, there exist vectors $\zeta^{(k)}_u = (\zeta^{(k)}_{u,1}, \zeta^{(k)}_{u,2}, \dots, \zeta^{(k)}_{u,k+1}) \in \mathbb{R}^{k+1}$, $u \in \{1/T, 2/T, \dots, 1\}$, satisfying $\|\zeta^{(k)}_u - \theta_{0,k}(u)\|_2 \le \|\hat\theta_{N,k}(u) - \theta_{0,k}(u)\|_2$ such that

$$\nabla\mathcal{L}^{\hat\mu}_{N,k}(\hat\theta_{N,k}(u), u) - \nabla\mathcal{L}^{\hat\mu}_{N,k}(\theta_{0,k}(u), u) = \nabla^2\mathcal{L}^{\hat\mu}_{N,k}(\zeta^{(k)}_u, u)\big(\hat\theta_{N,k}(u) - \theta_{0,k}(u)\big),$$

and the first term on the left-hand side vanishes due to (7.17). This yields

$$E_T - \nabla \mathcal{L}^{\mu}_{N,k}(\theta_{0,k}(u), u) \quad = \quad \nabla^2 \mathcal{L}^{\hat{\mu}}_{N,k}(\zeta^{(k)}_u, u)\big(\hat{\theta}_{N,k}(u) - \theta_{0,k}(u)\big),$$

where E_T denotes the difference between $\nabla \mathcal{L}^{\mu}_{N,k}(\theta_{0,k}(u), u)$ and $\nabla \mathcal{L}^{\hat{\mu}}_{N,k}(\theta_{0,k}(u), u)$, which is of order $\max_{t=1,\dots,T} \big|\mu(t/T) - \hat{\mu}(t/T)\big| O_p(k^{1/2} N^{\epsilon})$ by (7.16). It follows from

$$\nabla \mathcal{L}^{\mu}_{N,k}(\theta_k, u) \quad = \quad \frac{1}{4\pi} \int_{-\pi}^{\pi} \big[I^{\mu}_N(u, \lambda) - f_{\theta_k}(\lambda)\big] \nabla f^{-1}_{\theta_k}(\lambda)\, d\lambda$$

and Theorem 7.2 that $\max_{u \in \{1/T, \dots 1\}} \|\nabla \mathcal{L}^{\mu}_{N,k}(\theta_{0,k}(u), u)\|_2 = O_p(\sqrt{k} N^{-1/2+\epsilon/2})$ so it remains to show that

$$P(\nabla^2 \mathcal{L}^{\hat{\mu}}_{N,k}(\zeta^{(k)}_u, u)^{-1}$$

exists and that $\|\nabla^2 \mathcal{L}^{\hat{\mu}}_{N,k}(\zeta^{(k)}_u, u)^{-1}\|_{sp} \le Ck$ for all $u \in \{1/T, \dots, 1\}) \to 1$ for some positive constant C. This, however, follows with a Taylor expansion, (7.17), Theorem 7.2 and Assumption 4.1 (iv) for the corresponding expression with $\hat{\mu}$ replaced by μ. The more general case is then implied by the convergence-assumptions on $\hat{\mu}$. □

7.3. *Proof of Theorem 4.5 and Theorem 4.6*

We will show in Section 7.3.1 that under the null hypothesis H_0 the estimate

$$\max_{j=1,\dots,M} \big\|\hat{\theta}_{N,k}(u_j) - \theta_{0,k}(u_j)\big\|_2 \quad = \quad O_p(k^{3/2} N^{-1/2+\epsilon/2}) \tag{7.18}$$

is valid, while Theorem 4.3 and (4.16) imply

$$k^{3/2} \max_{j=1,\dots,M} \big\|\hat{\theta}_{N,k}(u_j) - \theta_{0,k}(u_j)\big\|_2 \quad = \quad o_p(1) \tag{7.19}$$

under the alternative H_1. As in the proof of Theorem 4.3 there exist vectors $\zeta^{(k)}_j = (\zeta^{(k)}_{j,1}, \zeta^{(k)}_{j,2}, \dots, \zeta^{(k)}_{j,k+1}) \in \mathbb{R}^{k+1}$, $j = 1, \dots, M$, satisfying

$$\|\zeta^{(k)}_j - \theta_{0,k}(u_j)\|_2 \le \|\hat{\theta}_{N,k}(u_j) - \theta_{0,k}(u_j)\|_2$$

such that

$$-\nabla \mathcal{L}^{\hat{\mu}}_{N,k}(\theta_{0,k}(u_j), u_j) \quad = \quad \nabla^2 \mathcal{L}^{\hat{\mu}}_{N,k}(\zeta^{(k)}_j, u_j)\big(\hat{\theta}_{N,k}(u_j) - \theta_{0,k}(u_j)\big)$$

holds because of Assumption 4.1 (ii) and (7.18) (under H_0) or (7.19) (under H_1). By rearranging and summing over every block, it follows that

$$\frac{1}{M} \sum_{j=1}^{M} \big(\hat{\theta}_{N,k}(u_j) - \theta_{0,k}(u_j)\big) \quad = \quad R_{0,T} - R_{1,T} - R_{2,T} - R_{3,T} - R_{4,T} \tag{7.20}$$

where

$$R_{0,T} \quad := \quad -\frac{1}{M}\sum_{j=1}^{M}\Gamma_k^{-1}(\theta_{0,k}(u_j))\nabla\mathcal{L}_{N,k}^{\mu}(\theta_{0,k}(u_j),u_j),$$

Γ_k^{-1} is defined in (4.5) and the terms $R_{i,T}(i=1\ldots,4)$ are given by

$$\begin{aligned}
R_{1,T} \quad &:= \quad \frac{1}{M}\sum_{j=1}^{M}\Gamma_k^{-1}(\theta_{0,k}(u_j))\big(\nabla\mathcal{L}_{N,k}^{\hat\mu}(\theta_{0,k}(u_j),u_j)-\nabla\mathcal{L}_{N,k}^{\mu}(\theta_{0,k}(u_j),u_j)\big),\\
R_{2,T} \quad &:= \quad \frac{1}{M}\sum_{j=1}^{M}\Gamma_k^{-1}(\theta_{0,k}(u_j))\big(\nabla^2\mathcal{L}_{N,k}^{\hat\mu}(\zeta_j^{(k)},u_j)\\
&\qquad\qquad -\nabla^2\mathcal{L}_{N,k}^{\mu}(\zeta_j^{(k)},u_j)\big)\big(\hat\theta_{N,k}(u_j)-\theta_{0,k}(u_j)\big),\\
R_{3,T} \quad &:= \quad \frac{1}{M}\sum_{j=1}^{M}\Gamma_k^{-1}(\theta_{0,k}(u_j))\big(\nabla^2\mathcal{L}_{N,k}^{\mu}(\zeta_j^{(k)},u_j)\\
&\qquad\qquad -\nabla^2\mathcal{L}_{N,k}^{\mu}(\theta_{0,k}(u_j),u_j)\big)\big(\hat\theta_{N,k}(u_j)-\theta_{0,k}(u_j)\big),\\
R_{4,T} \quad &:= \quad \frac{1}{M}\sum_{j=1}^{M}\Gamma_k^{-1}(\theta_{0,k}(u_j))\big(\nabla^2\mathcal{L}_{N,k}^{\mu}(\theta_0(u_j),u_j)\\
&\qquad\qquad -\Gamma_k(\theta_{0,k}(u_j))\big)\big(\hat\theta_{N,k}(u_j)-\theta_{0,k}(u_j)\big).
\end{aligned}$$

We obtain for the first summand in (7.20)

$$R_{0,T}=-\frac{1}{M}\sum_{j=1}^{M}\frac{1}{4\pi}\int_{-\pi}^{\pi}\big[I_N^{\mu}(u_j,\lambda)-f_{\theta_{0,k}(u_j)}(\lambda)\big]\Gamma_k^{-1}(\theta_{0,k}(u_j))\nabla f_{\theta_{0,k}(u_j)}^{-1}(\lambda)\,d\lambda$$

and with the notation $\phi_T(u_j,\lambda)=1/(4\pi)[\Gamma_k^{-1}(\theta_{0,k}(u_j))\nabla f_{\theta_{0,k}(u_j)}^{-1}(\lambda)]_1$, it is easy to see that Assumption 4.1 (i)–(iv) imply the conditions of Theorem 7.1 b) with $g(k)=k^2$. Moreover, observing the definition of V_T and W_T in Theorem 7.1 and 4.5, (4.6) yields $V_T/W_T\to 1$. Consequently, under the assumptions of Theorem 4.5 it follows (observing (4.9) and the growth conditions on N, T)

$$\frac{\sqrt{T}}{M}\sum_{j=1}^{M}\big[\Gamma_k^{-1}(\theta_{0,k}(u_j))\nabla\mathcal{L}_N^{\mu}(\theta_{0,k}(u_j),u_j)\big]_1/\sqrt{W_T}\xrightarrow{\mathcal{D}}\mathcal{N}(0,1).$$

Since $d_0(u)$ is the first element of the vector $\theta_{0,k}(u)$, Theorem 4.5 is a consequence of the fact $\frac{1}{M}\sum_{j=1}^{M}d_0(u_j)=F+O(M^{-2})$ [this can be proved by a second order Taylor expansion] if we are able to show that

$$R_{i,T}=o_p(T^{-1/2});\quad i=1,\ldots,4.$$

Analogously, Theorem 4.6 follows from (7.4) and (7.5) if the estimates

$$R_{i,T}=o_p(1)\quad i=1,\ldots,4.$$

can be established. It can be shown analogously to the proof of Theorem 3.6 in Dahlhaus (1997), that, under assumptions (4.13) – (4.14), both terms $R_{1,T}$ and

$R_{2,T}$ are of order $O_p(k^2 N^{-\epsilon} T^{-1/2} + k^2 N^{\epsilon-1})$, while, under assumption (4.16), the order is $o_p(1)$ [see the proof of (7.23) and (7.15), respectively, for more details]. Therefore it only remains to consider the quantities $R_{3,T}$ and $R_{4,T}$. For this purpose note that

$$\begin{aligned}\nabla^2 \mathcal{L}^\mu_{N,k}(\theta_k(u_j), u_j) &= \frac{1}{4\pi}\int_{-\pi}^{\pi} \big[I^\mu_N(u_j,\lambda) - f_{\theta_k(u_j)}(\lambda)\big]\nabla^2 f^{-1}_{\theta_k(u_j)}(\lambda)\, d\lambda \\ &\qquad + \Gamma_k(\theta_k(u_j)) \qquad (7.21)\\ \nabla^3 \mathcal{L}^\mu_{N,k}(\theta_k(u_j), u_j) &= \frac{1}{4\pi}\int_{-\pi}^{\pi} \big[I^\mu_N(u_j,\lambda) - f_{\theta_k(u_j)}(\lambda)\big] \\ &\qquad \times \left[\frac{\partial^3 f^{-1}_{\theta_k(u_j)}(\lambda)}{\partial\theta_{j,t}\partial\theta_{j,s}\partial\theta_{j,r}}\right]_{r,s,t=1,\dots,k+1} d\lambda \\ &\quad -\frac{1}{4\pi}\int_{-\pi}^{\pi}\left[\frac{\partial f_{\theta_k(u_j)}(\lambda)}{\partial\theta_{j,t}}\frac{\partial^2 f^{-1}_{\theta_k(u_j)}(\lambda)}{\partial\theta_{j,s}\partial\theta_{j,r}}\right]_{r,s,t=1,\dots,k+1} d\lambda \qquad (7.22)\\ &\quad +\frac{1}{4\pi}\int_{-\pi}^{\pi}\left[\frac{\partial}{\partial\theta_{j,t}}\left(\frac{\partial f_{\theta_k(u_j)}(\lambda)}{\partial\theta_{j,s}}\frac{1}{f^2_{\theta_k(u_j)}(\lambda)}\frac{\partial f_{\theta_k(u_j)}(\lambda)}{\partial\theta_{j,r}}\right)\right]_{r,s,t=1,\dots,k+1} d\lambda,\end{aligned}$$

where we used the notation $(\theta_{j,1}, \theta_{j,2}, \dots, \theta_{j,k+1}) := (d(u_j), a_1(u_j), \dots, a_k(u_j))$. For the term $R_{3,T}$ we obtain with the well-known inequality $\|Ax\|_2 \le \|A\|_{sp}\|x\|_2$

$$\begin{aligned}\|R_{3,T}\|_2 &\le \max_{\theta_k\in\Theta_{R,k}} \big\|\Gamma_k^{-1}(\theta_k)\big\|_{sp} \\ &\quad\times \frac{1}{M}\sum_{j=1}^{M}\big\|\nabla^2\mathcal{L}^\mu_{N,k}(\zeta_j^{(k)}, u_j) - \nabla^2\mathcal{L}^\mu_{N,k}(\theta_{0,k}(u_j), u_j)\big\|_{sp}\big\|\hat\theta_{N,k}(u_j) - \theta_{0,k}(u_j)\big\|_2.\end{aligned}$$

By the mean value theorem there exist vectors $\tilde\zeta_j^{(k)} \in \mathbb{R}^k$ such that

$$\begin{aligned}&\big\|\nabla^2\mathcal{L}^\mu_{N,k}(\zeta_j^k, u_j) - \nabla^2\mathcal{L}^\mu_{N,k}(\theta_0(u_j), u_j)\big\|_{sp} \\ \le\ & k \max_{r,s=1,\dots,k}\big|\big[\nabla^2\mathcal{L}^\mu_{N,k}(\zeta_j^{(k)}, u_j) - \nabla^2\mathcal{L}^\mu_{N,k}(\theta_{0,k}(u_j), u_j)\big]_{r,s}\big| \\ =\ & k \max_{r,s=1,\dots,k}\big|\nabla\big[\nabla^2\mathcal{L}^\mu_{N,k}(\tilde\zeta_j^{(k)}, u_j)\big]_{r,s}\big(\zeta_j^{(k)} - \theta_{0,k}(u_j)\big)\big| \\ \le\ & k \max_{r,s=1,\dots,k}\big\|\nabla\big[\nabla^2\mathcal{L}^\mu_{N,k}(\tilde\zeta_j^{(k)}, u_j)\big]_{r,s}\big\|_2\big\|\zeta_j^{(k)} - \theta_{0,k}(u_j)\big\|_2 \\ \le\ & k\big\|\hat\theta_{N,k}(u_j) - \theta_{0,k}(u_j)\big\|_2 \sup_{\substack{\theta_k\in\Theta_{R,k}\\ r,s=1,\dots,k}}\big\|\nabla\big[\nabla^2\mathcal{L}^\mu_{N,k}(\theta_k, u_j)\big]_{r,s}\big\|_2,\end{aligned}$$

where $\|\tilde\zeta_j^{(k)} - \theta_{0,k}(u_j)\|_2 \le \|\zeta_j^{(k)} - \theta_{0,k}(u_j)\|_2$ for every $j = 1, ..., M$. Therefore, we obtain

$$\begin{aligned}\|R_{3,T}\|_2 &\le k \max_{j=1,\dots,M}\big\|\hat\theta_{N,k}(u_j) - \theta_{0,k}(u_j)\big\|_2^2 \sup_{\theta_k\in\Theta_{R,k}}\big\|\Gamma_k^{-1}(\theta_k)\big\|_{sp} \\ &\qquad\times \sup_{\substack{\theta_k\in\Theta_{R,k}; j=1,\dots,M\\ r,s=1,\dots k}}\big\|\nabla\big[\nabla^2\mathcal{L}^\mu_N(\theta_k, u_j)\big]_{r,s}\big\|_2\end{aligned}$$

$$\leq \quad kC \max_{j=1,\dots,M} \left\|\hat{\theta}_{N,k}(u_j) - \theta_{0,k}(u_j)\right\|_2^2 \sup_{\theta_k \in \Theta_{R,k}} \left\|\Gamma_k^{-1}(\theta_k)\right\|_{sp}$$

$$\Big(k \cdot \sup_{\substack{\theta_k \in \Theta_{R,k}; j=1,\dots,M \\ r,s,t=1,\dots,k}} \Big| \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[I_N^\mu(u_j,\lambda) - f_{\theta_k}(\lambda)\right] \frac{\partial^3 f_{\theta_k}^{-1}(\lambda)}{\partial\theta_{j,t}\partial\theta_{j,s}\partial\theta_{j,r}}\, d\lambda \Big| + k\Big),$$

where, in the last inequality, we have used the fact that the second and third term in (7.22) are bounded by a constant [this follows directly from Assumption 4.1]. Before we investigate the order of this expression, we derive a similar bound for the term $R_{4,T}$. Observing (7.22) we obtain

$$\begin{aligned}
\|R_{4,T}\|_2 &\leq \max_{j=1,\dots,M} \left\|\hat{\theta}_{N,k}(u_j) - \theta_{0,k}(u_j)\right\|_2 \sup_{\theta_k \in \Theta_{R,k}} \left\|\Gamma_k^{-1}(\theta_k)\right\|_{sp} \\
&\quad \times \max_{j=1,\dots,M} \left\|\nabla^2 \mathcal{L}_{N,k}^\mu(\theta_{0,k}(u_j), u_j) - \Gamma_k(\theta_{0,k}(u_j))\right\|_{sp} \\
&= \max_{j=1,\dots,M} \left\|\hat{\theta}_{N,k}(u_j) - \theta_{0,k}(u_j)\right\|_2 \sup_{\theta_k \in \Theta_{R,k}} \left\|\Gamma_k^{-1}(\theta_k)\right\|_{sp} \\
&\quad \times \max_{j=1,\dots,M} \Big\| \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[I_N^\mu(u_j,\lambda) - f_{\theta_{0,k}(u_j)}(\lambda)\right] \nabla^2 f_{\theta_{0,k}(u_j)}^{-1}(\lambda)\, d\lambda \Big\|_{sp} \\
&\leq k \max_{j=1,\dots,M} \left\|\hat{\theta}_{N,k}(u_j) - \theta_{0,k}(u_j)\right\|_2 \sup_{\theta_k \in \Theta_{R,k}} \left\|\Gamma_k^{-1}(\theta_k)\right\|_{sp} \\
&\quad \times \max_{j=1,\dots,M} \max_{r,s=1,\dots,k} \Big| \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[I_N^\mu(u_j,\lambda) - f_{\theta_{0,k}(u_j)}(\lambda)\right] \\
&\quad \times \frac{\partial^2 f_{\theta_{0,k}(u_j)}^{-1}(\lambda)}{\partial\theta_{j,s}\partial\theta_{j,r}}\, d\lambda \Big|.
\end{aligned}$$

If we show

$$\max_{j=1,\dots,M} \sup_{\substack{\theta_k \in \Theta_{R,k} \\ r,s,t=1,\dots,k}} \Big| \int_{-\pi}^{\pi} \left[I_N^\mu(u_j,\lambda) - f_{\theta_k}(\lambda)\right] \frac{\partial^3 f_{\theta_k}^{-1}(\lambda)}{\partial\theta_{j,t}\partial\theta_{j,s}\partial\theta_{j,r}}\, d\lambda \Big| = O_p(1),$$

$$\begin{aligned}
&\max_{j=1,\dots,M} \max_{r,s=1,\dots,k} \Big| \int_{-\pi}^{\pi} \left[I_N^\mu(u_j,\lambda) - f_{\theta_{0,k}(u_j)}(\lambda)\right] \frac{\partial^2 f_{\theta_{0,k}(u_j)}^{-1}(\lambda)}{\partial\theta_{j,s}\partial\theta_{j,r}}\, d\lambda \Big| \\
&\quad = O_p\Big(\frac{1}{N^{1/2-\epsilon}}\Big),
\end{aligned}$$

it follows with Assumption 4.1 (iv) in combination with (7.18) (under H_0) and (7.19) (under H_1) that the terms $R_{3,T}$ and $R_{4,T}$ are of order $o_p(T^{-1/2})$ (under H_0) and $o_p(1)$ (under H_1). These two claims, however are a direct consequence of Theorem 7.2 and (4.9). □

7.3.1. Proof of (7.18)

With the same arguments as in the proof of Theorem 3.6 in Dahlhaus (1997) we obtain

$$\max_{j=1,\dots,M} \left|\mathcal{L}_{N,k}^{\hat{\mu}}(\theta_k, u_j) - \mathcal{L}_{N,k}^\mu(\theta_k, u_j)\right| \leq \Pi_{1,T} + \Pi_{2,T},$$

where

$$\begin{aligned}\Pi_{1,T} &= C \max_{t=1,\dots,T} \max_{q=1,\dots,N} \sup_{\theta_k \in \Theta_{R,k}} \Big| \int_{-\pi}^{\pi} d_N^{X-\mu}(t/T,\lambda) f_{\theta_k}^{-1}(\lambda) \sum_{s=0}^{q-1} e^{is\lambda}\, d\lambda \Big| \\ &\quad \times \Big(\max_{t=1,\dots,T} \Big| \Big\{ \mu\Big(\frac{t-1}{T}\Big) - \hat\mu\Big(\frac{t-1}{T}\Big) \Big\} - \Big\{ \mu\Big(\frac{t}{T}\Big) - \hat\mu\Big(\frac{t}{T}\Big) \Big\} \Big| \\ &\qquad + \max_{t=1,\dots,T} \big| \mu(t/T) - \hat\mu(t/T) \big| / N \Big) \\ \Pi_{2,T} &= C N^{\epsilon} \max_{t=1,\dots,T} \big| \mu(t/T) - \hat\mu(t/T) \big|^2 \end{aligned}$$

and C denotes a positive constant. By proceeding as in the proof of Theorem 7.2 one obtains

$$\max_{t=1,\dots,T} \max_{q=1,\dots,N} \sup_{\theta_k \in \Theta_{R,k}} \Big| \int_{-\pi}^{\pi} d_N^{X-\mu}(t/T,\lambda) f_{\theta_k}^{-1}(\lambda) \sum_{s=0}^{q-1} e^{is\lambda}\, d\lambda \Big| = o(N^{1/2+\epsilon/2}),$$

which implies (observing the assumptions (4.13) and (4.14))

$$\begin{aligned}\max_{j=1,\dots,M} \sup_{\theta_k \in \Theta_{R,k}} \big| \mathcal{L}_{N,k}^{\hat\mu}(\theta_k, u_j) - \mathcal{L}_{N,k}^{\mu}(\theta_k, u_j) \big| &= O_p(N^{-\epsilon} T^{-1/2} + N^{\epsilon-1}) \\ &= o_P(N^{-1/2+\epsilon/2} k^{1/2}) \qquad (7.23)\end{aligned}$$

under H_0. Analogously we obtain

$$\begin{aligned} &\max_{j=1,\dots,M} \sup_{\theta_k \in \Theta_{R,k}} \big\| \nabla \mathcal{L}_{N,k}^{\hat\mu}(\theta_k, u_j) - \nabla \mathcal{L}_{N,k}^{\mu}(\theta_k, u_j) \big\|_2 \\ = &\; O_p(k^{1/2} N^{-\epsilon} T^{-1/2} + k^{1/2} N^{\epsilon-1}) = o_P(N^{-1/2+\epsilon/2} k^{1/2}) \qquad (7.24)\end{aligned}$$

under the null hypothesis. By using (7.23) and (7.24) instead of (7.15) and (7.16), assertion (7.18) follows by the same arguments as given in the proof of Theorem 4.3. □

7.4. *Proof of Theorem 4.10*

A second order Taylor expansion yields

$$\begin{aligned}\mathbb{E}(\hat\mu_L(t/T)) &= \mu(t/T) + \frac{\mu'(t/T)}{L} \sum_{p=0}^{L-1} (-L/2 + 1 + p)/T + O(L^2/T^2) \\ &= \mu(t/T) + O(1/T + L^2/T^2). \qquad (7.25)\end{aligned}$$

For $t_i \in \{1, ..., T\}$ the cumulants of order $l \geq 2$

$$\mathrm{cum}(\hat\mu_L(t_1/T), \hat\mu_L(t_2/T), ..., \hat\mu_L(t_l/T)) = \frac{1}{L^l} \sum_{p_1,\dots,p_l=0}^{L-1} \sum_{m_1,\dots,m_l=0}^{\infty}$$

$$\psi_{t_1,T,m_1}\cdots\psi_{t_l,T,m_l}\mathrm{cum}(Z_{p_1-m_1},...,Z_{p_l-m_l})$$

are bounded by

$$\frac{C}{L^l}\sum_{p_1=0}^{L-1}\sum_{\substack{m_1,...,m_l=0\\|m_i-m_{i+1}|\le L}}^{\infty}\frac{1}{(I(m_1\cdots m_l))^{1-D}}\le C^lL^{1-l(1-D)},$$

where we used the independence of the innovations, (2.3) and (2.4) and the last inequality follows by replacing the sums by its corresponding approximating integrals and holds for some positive constant C (which is independent of l and may vary in the following arguments). This yields that $\hat{\mu}_L(t/T)$ estimates its true counterpart at a pointwise rate of $L^{1/2-D}$ and we now continue by showing stochastic equicontinuity. The expansion (7.25) and the bound $C^lL^{1-l(1-D)}$ for the l-th cumulant ($l\ge 2$) of $\hat{\mu}_L$ yield $\mathrm{cum}_l(L^{1/2-D-\alpha/2}(\hat{\mu}_L(t_1/T)-\hat{\mu}_L(t_2/T)))\le(2C)^lL^{-l\alpha/2}$ for all $t_i\in\{1,...,T\}$ and every $\alpha>0$, from which we get

$$\mathbb{E}(L^{l(1/2-D-\alpha)}(\hat{\mu}_L(t_1/T)-\hat{\mu}_L(t_2/T))^l)\le(2l)!C^lL^{-l\alpha/2}$$

for all even $l\in\mathbb{N}$ and $t_i\in\{1,\dots,T\}$ [see the proof of Lemma 2.3 in Dahlhaus (1988) for more details]. By considering the order of the bias (7.25) this yields

$$L^{1/2-D-\alpha}\max_{t=1,\dots,T}\big|\mu(t/T)-\hat{\mu}_L(t/T)\big|=o_p(1),\qquad\text{for every }\alpha>0,$$

as in the proof of Theorem 7.2. Consequently (4.13) [under the conditions of part a)] and (4.16) [under the conditions of part b)] follow. So it remains to show (4.14) in the case $D=0$. For this purpose we define

$$\Delta(t/T)=\Big\{\mu\Big(\frac{t-1}{T}\Big)-\hat{\mu}_L\Big(\frac{t-1}{T}\Big)\Big\}-\Big\{\mu\Big(\frac{t}{T}\Big)-\hat{\mu}_L\Big(\frac{t}{T}\Big)\Big\},$$

and from (7.25) we obtain $\mathbb{E}(\Delta(t/T))=O(T^{-1}+L^2/T^2)$. A simple calculation reveals $\mathrm{cum}(\Delta(t_1/T),\Delta(t_2/T))=O(L^{-1}T^{-1})$ (where the estimate is independent of t_i) and with the Gaussianity of the innovations we get

$$\mathrm{cum}(\Delta(t_1/T),...,\Delta(t_l/T))=0$$

for $l\ge 3$. This yields, as above,

$$L^{1/2-\alpha}T^{1/2}\max_{t=1,\dots,T}|\Delta(t/T)|=o_p(1)$$

for every $\alpha>0$, and completes the proof of Theorem 4.10. □

8. Auxiliary results

Finally, we state some lemmas which were employed in the above proofs.

Lemma 8.1. *Suppose it is $\mu,\nu,a,b\in\mathbb{R}$. Then there exists a constant $C\in\mathbb{R}$ such that the following holds:*

a) If $\mu, \nu > 0$ and $b > a$, then

$$\sum_{\substack{p=0\\p-a\geq 1\\-p+b\geq 1}}^{N-1} \frac{1}{(p-a)^{1-\mu}} \frac{1}{(b-p)^{1-\nu}} \leq \sum_{p=1+a}^{b-1} \frac{1}{(p-a)^{1-\mu}} \frac{1}{(b-p)^{1-\nu}} \leq \frac{C}{(b-a)^{1-\mu-\nu}}. \tag{7.1}$$

b) If $0 < \mu, \nu$ and $0 < 1-\mu-\nu$, then it follows for $|a+b| > 0$

$$\sum_{\substack{p=1\\p+b\geq 1\\p-a\geq 1}}^{N-1} \frac{1}{(p+b)^{1-\mu}} \frac{1}{(p-a)^{1-\nu}} \leq \sum_{\substack{p=1\\p+b\geq 1\\p-a\geq 1}}^{\infty} \frac{1}{(p+b)^{1-\mu}} \frac{1}{(p-a)^{1-\nu}} \leq \frac{C}{|a+b|^{1-\mu-\nu}}. \tag{7.2}$$

c) If $0 < \nu < 1-\mu$ and $y, z \geq 1$, then

$$\sum_{p=1+y}^{\infty} \frac{\log(p)}{p^{1-\mu}} \frac{1}{(p-y)^{1-\nu}} \leq \frac{C\log(y)}{y^{1-\mu-\nu}},$$

$$\sum_{p=1}^{\infty} \frac{\log(p+z)}{(p+z)^{1-\mu}} \frac{1}{p^{1-\nu}} \leq \frac{C\log(z)}{z^{1-\mu-\nu}}.$$

Proof: The proof can be found in Sen et al. (2016). □

Lemma 8.2. *For every $T \in \mathbb{N}$, let $\eta_T : [-\pi, \pi] \mapsto \mathbb{R}$ be a symmetric and twice continuously differentiable function such that $\eta_T = O(|\lambda|^\alpha)$ for some $\alpha \in (-1, 1)$ as $|\lambda| \to 0$ (where the constant in the $O(\cdot)$ term is independent of T). Then, for $|h| \to \infty$, we have*

$$\int_{-\pi}^{\pi} \eta_T(\lambda) e^{ih\lambda}\, d\lambda = O\Big(\frac{1}{|h|^{1+\alpha}}\Big)$$

uniformly in T.

Proof: The assertion follows from Lemma 4 and Lemma 5 in Fox and Taqqu (1986). □

Lemma 8.3. *If Assumption 2.1 holds, then*

(i)

$$\frac{1}{N^2} \frac{1}{M^2} \sum_{j_1=1}^{M} \sum_{q,r=0}^{N-1} \sum_{l,m,n,o=0}^{\infty} \sum_{\substack{j_2=1\\N\leq|r+l-n+(j_2-j_1)N|\\0\leq q+o-m+(j_1-j_2)N\leq N-1\\|l-n+(j_2-j_1)N|\leq N-1\\|o-m+(j_1-j_2)N|\leq N-1}}^{M} \psi_l(u_{j_1})\psi_m(u_{j_1})\psi_n(u_{j_2})\psi_o(u_{j_2})$$

$$\int_{-\pi}^{\pi} \phi_T(u_{j_1},\lambda_1) e^{-i(r-q+l-n+(j_2-j_1)N)\lambda_1}\,d\lambda_1 \times \int_{-\pi}^{\pi} \phi_T(u_{j_2},\lambda_2) e^{-i(r-q+m-o+(j_2-j_1)N)\lambda_2}\,d\lambda_2 = O\Big(\frac{1}{T}\frac{g^2(k)}{N^{1-2D-2\epsilon}}\Big)$$

(ii)

$$\frac{1}{N^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{q,r=0}^{N-1}\sum_{l,m,n,o=0}^{\infty}\sum_{\substack{j_2=1\\ -(N-1)\le r+l-n+(j_2-j_1)N\le -1\\ 0\le q+o-m+(j_1-j_2)N\le N-1\\ |l-n+(j_2-j_1)N|\le N-1\\ |o-m+(j_1-j_2)N|\le N-1}}^{M} \psi_l(u_{j_1})\psi_m(u_{j_1})\psi_n(u_{j_2})\psi_o(u_{j_2})$$

$$\int_{-\pi}^{\pi} \phi_T(u_{j_1},\lambda_1) e^{-i(r-q+l-n+(j_2-j_1)N)\lambda_1}\,d\lambda_1 \times \int_{-\pi}^{\pi} \phi_T(u_{j_2},\lambda_2) e^{-i(r-q+m-o+(j_2-j_1)N)\lambda_2}\,d\lambda_2 = O\Big(\frac{1}{T}\frac{g^2(k)}{N^{1-2D-2\epsilon}}\Big)$$

(iii)

$$\frac{1}{N^2}\frac{1}{M^2}\sum_{j_2=1}^{M}\sum_{q=0}^{N-1}\sum_{r=-(N-1)}^{-1}\sum_{l,m,n,o=0}^{\infty}\sum_{\substack{j_1=1\\ 0\le q+o-m+(j_1-j_2)N\le N-1\\ |l-n+(j_2-j_1)N|\le N-1\\ |o-m+(j_1-j_2)N|\le N-1}}^{M} \psi_l(u_{j_1})\psi_m(u_{j_1})\psi_n(u_{j_2})\psi_o(u_{j_2})$$

$$\int_{-\pi}^{\pi} \phi_T(u_{j_1},\lambda_1) e^{-i(r-q+l-n+(j_2-j_1)N)\lambda_1}\,d\lambda_1 \times \int_{-\pi}^{\pi} \phi_T(u_{j_2},\lambda_2) e^{-i(r-q+m-o+(j_2-j_1)N)\lambda_2}\,d\lambda_2 = O\Big(\frac{1}{T}\frac{g^2(k)}{N^{1-2D-2\epsilon}}\Big)$$

(iv)

$$\frac{1}{N^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{r,q=0}^{N-1}\sum_{l,m,n,o=0}^{\infty}\sum_{\substack{j_2=1\\ 0\le q+o-m+(j_1-j_2)N\le N-1\\ N\le |l-n+(j_2-j_1)N|\\ |o-m+(j_1-j_2)N|\le N-1}}^{M} \psi_l(u_{j_1})\psi_m(u_{j_1})\psi_n(u_{j_2})\psi_o(u_{j_2})$$

$$\int_{-\pi}^{\pi} \phi_T(u_{j_1},\lambda_1)\phi_T(u_{j_2},\lambda_1) e^{-i(r-q+l-n+(j_2-j_1)N)\lambda_1}\,d\lambda_1 \times \int_{-\pi}^{\pi} e^{-i(r-q+m-o+(j_2-j_1)N)\lambda_2}\,d\lambda_2 = O\Big(\frac{1}{T}\frac{g^2(k)}{N^{1-2D-2\epsilon}}\Big)$$

(v)

$$\frac{1}{N^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{r,q=0}^{N-1}\sum_{l,m,n,o=0}^{\infty}\sum_{\substack{j_2=1\\ |j_1-j_2|\ge 1\\ 0\le q+o-m+(j_1-j_2)N\le N-1\\ |o-m+(j_1-j_2)N|\le N-1}}^{M} \psi_l(u_{j_1})\psi_m(u_{j_1})\psi_n(u_{j_2})\psi_o(u_{j_2})$$

$$\int_{-\pi}^{\pi} \phi_T(u_{j_1},\lambda_1)\phi_T(u_{j_2},\lambda_1)e^{-i(r-q+l-n+(j_2-j_1)N)\lambda_1}\,d\lambda_1 \times \int_{-\pi}^{\pi} e^{-i(r-q+m-o+(j_2-j_1)N)\lambda_2}\,d\lambda_2 = O\Big(\frac{1}{T}\frac{g^2(k)}{N^{1-2D-2\epsilon}}\Big)$$

(vi)

$$\frac{1}{N^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{q=0}^{N-1}\sum_{\substack{r\in\mathbb{Z}\\|r|\geq N}}\sum_{l,m,n,o=0}^{\infty}\sum_{\substack{j_2=1\\0\leq q+o-m+(j_1-j_2)N\leq N-1\\|l-n+(j_2-j_1)N|\leq N-1\\|o-m+(j_1-j_2)N|\leq N-1}}^{\infty}\psi_l(u_{j_1})\psi_m(u_{j_1})\psi_n(u_{j_2})\psi_o(u_{j_2})$$
$$\int_{-\pi}^{\pi}\phi_T(u_{j_1},\lambda_1)e^{-i(r-q+l-n+(j_2-j_1)N)\lambda_1}$$
$$\times\int_{-\pi}^{\pi}\Big[\phi_T(u_{j_2},\lambda_2)-\phi_T(u_{j_2},\lambda_1)\Big]e^{-i(r-q+m-o+(j_2-j_1)N)\lambda_2}\,d\lambda_2\,d\lambda_1$$
$$=O\Big(\frac{1}{T}\frac{g^2(k)}{N^{1-2D-2\epsilon}}\Big)$$

Proof: Without loss of generality we restrict ourselves to a proof of part (i) and (v) and note that all other claims are proven by using the same arguments.

Proof of (i): We use (2.4), (7.1) and Lemma 8.2 to bound the term in (i) (up to a constant) through

$$\frac{g^2(k)}{N^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{q,r=0}^{N-1}\sum_{l,m,n,o=1}^{\infty}\sum_{\substack{j_2=1\\N\leq|r+l-n+(j_2-j_1)N|\\0\leq q+o-m+(j_1-j_2)N\leq N-1\\|l-n+(j_2-j_1)N|\leq N-1\\|o-m+(j_1-j_2)N|\leq N-1\\1\leq|r-q+m-o+(j_2-j_1)N|}}^{M}\frac{1}{l^{1-d_0(u_{j_1})}}\frac{1}{m^{1-d_0(u_{j_1})}}\frac{1}{n^{1-D}}\frac{1}{o^{1-D}}$$
$$\frac{1}{|r-q+l-n+(j_2-j_1)N|^{1+2d_0(u_{j_1})-\epsilon}}\frac{1}{|r-q+m-o+(j_2-j_1)N|^{1-\epsilon}}.$$

If the variables j_1,o and m are fixed, it follows with the constraint $0\leq q+o-m+(j_1-j_2)N\leq N-1$ that there are at most two possible values for j_2 such that the resulting term is non vanishing. We now discuss for which combinations of j_1 and j_2 the above expression is maximized and then restrict ourselves to the resulting pair (j_1,j_2).

If j_1 and j_2 are given, the variables l,m,n,o can only be chosen such that $|l-n+(j_2-j_1)N|\leq N-1$ and $|o-m+(j_1-j_2)N|\leq N-1$ are fulfilled. Therefore, the possible values of the fractions $(|r-q+l-n+(j_2-j_1)N|)^{-1}(|r-q+m-o+(j_2-j_1)N|)^{-1}$ are the same for any combination of j_1 and j_2. Consequently, in order to maximize the term above we need to maximize $l^{-1d_0(u_{j_1})}m^{-1+d_0(u_{j_1})}n^{-1+D}o^{-1+D}$, which is achieved by the choice

$j_1 = j_2$ [since then l, m, n, o can be jointly taken as small as possible due to the constraints $|l - n + (j_2 - j_1)N| \leq N - 1$ and $|o - m + (j_1 - j_2)N| \leq N - 1$]. Hence we can bound that above expression (up to a constant) by

$$\frac{g^2(k)}{N^2} \frac{1}{M^2} \sum_{j_1=1}^{M} \sum_{q,r=0}^{N-1} \sum_{\substack{l,m,n,o=1 \\ N \leq |r+l-n| \\ |l-n| \leq N-1 \\ |o-m| \leq N-1 \\ 1 \leq |r-q+m-o|}}^{\infty} \frac{1}{l^{1-d_0(u_{j_1})}} \frac{1}{m^{1-d_0(u_{j_1})}} \frac{1}{n^{1-D}} \frac{1}{o^{1-D}}$$
$$\times \frac{1}{|r-q+l-n|^{1+2d_0(u_{j_1})-\epsilon}} \frac{1}{|r-q+m-o|^{1-\epsilon}}.$$

By setting $g := r + l - n$ and $s := q + o - m$ this term can be written as

$$\frac{g^2(k)}{N^2} \frac{1}{M^2} \sum_{j_1=1}^{M} \sum_{\substack{q,r,s=0 \\ 1 \leq |r-s|}}^{N-1} \sum_{\substack{g \in \mathbb{Z} \\ |g| \geq N}} \sum_{\substack{m,n=1 \\ 1 \leq g-r+n \\ 1 \leq s-q+m \\ |g-r| \leq N-1}}^{\infty} \frac{1}{(g-r+n)^{1-d_0(u_{j_1})}} \frac{1}{m^{1-d_0(u_{j_1})}} \frac{1}{n^{1-D}}$$
$$\times \frac{1}{(s-q+m)^{1-D}} \frac{1}{|g-q|^{1+2d_0(u_{j_1})-\epsilon}} \frac{1}{|r-s|^{1-\epsilon}}$$

Through an repeated application of (7.1) and (7.3) the claim now follows.

Proof of (v): By setting

$$f(u_{j_1}, u_{j_2}, \lambda) := \frac{1}{2\pi} \sum_{l,n=0}^{\infty} \psi_l(u_{j_1}) \psi_n(u_{j_2}) e^{-i(l-n)\lambda}.$$

we can write the term in (v) as

$$\frac{2\pi}{N^2} \frac{1}{M^2} \sum_{j_1=1}^{M} \sum_{r,q=0}^{N-1} \sum_{m,o=0}^{\infty} \sum_{\substack{j_2=1 \\ |j_1-j_2| \geq 1 \\ 0 \leq q+o-m+(j_1-j_2)N \leq N-1 \\ |o-m+(j_1-j_2)N| \leq N-1}}^{M} \psi_m(u_{j_1}) \psi_o(u_{j_2})$$
$$\times \int_{-\pi}^{\pi} \phi_T(u_{j_1}, \lambda_1) \phi_T(u_{j_2}, \lambda_1) f(u_{j_1}, u_{j_2}, \lambda_1) e^{-i(r-q+(j_2-j_1)N)\lambda_1} \, d\lambda_1$$
$$\times \int_{-\pi}^{\pi} e^{-i(r-q+m-o+(j_2-j_1)N)\lambda_2} \, d\lambda_2.$$

and by integrating over λ_2 this is the same as

$$\frac{4\pi^2}{N^2} \frac{1}{M^2} \sum_{j_1=1}^{M} \sum_{q=0}^{N-1} \sum_{m,o=0}^{\infty} \sum_{\substack{j_2=1 \\ |j_1-j_2| \geq 1 \\ 0 \leq q+o-m+(j_1-j_2)N \leq N-1 \\ |o-m+(j_1-j_2)N| \leq N-1}}^{M} \psi_m(u_{j_1}) \psi_o(u_{j_2})$$

$$\times \int_{-\pi}^{\pi} \phi_T(u_{j_1}, \lambda_1)\phi_T(u_{j_2}, \lambda_1) f(u_{j_1}, u_{j_2}, \lambda_1) e^{-i(o-m)\lambda_1}\, d\lambda_1.$$

By (7.1) and Lemma 8.2 this sum can be bounded by

$$\frac{Cg^2(k)}{N^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{q=0}^{N-1}\sum_{m,o=1}^{\infty}\sum_{\substack{j_2=1\\ |j_1-j_2|\geq 1\\ 0\leq q+o-m+(j_1-j_2)N\leq N-1\\ |o-m+(j_1-j_2)N|\leq N-1}}^{M} \frac{1}{m^{1-d_0(u_{j_1})}}\frac{1}{o^{1-d_0(u_{j_1})}} \times \frac{1}{|o-m|^{1+d_0(u_{j_1})+d_0(u_{j_2})-2\epsilon}}$$

$$\leq \frac{Cg^2(k)}{N^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{q=0}^{N-1}\sum_{m,o=1}^{\infty}\sum_{\substack{j_2=1\\ |j_1-j_2|\geq 1\\ 0\leq q+o-m+(j_1-j_2)N\leq N-1\\ |o-m+(j_1-j_2)N|\leq N-1}}^{M} \frac{1}{m^{1-D}}\frac{1}{o^{1-D}}\frac{1}{|o-m|^{1-2\epsilon}}.$$

As in the proof of (i) we can argue that there are at most two possible values for j_2 if o, m and j_1 are chosen and that the expression is maximized for $|j_1-j_2| = 1$. Therefore we can bound the above expression up to a constant through

$$\frac{g^2(k)}{N^2}\frac{1}{M}\sum_{\kappa\in\{-1,1\}}\sum_{q=0}^{N-1}\sum_{\substack{m,o=1\\ 0\leq q+o-m+\kappa N\leq N-1\\ |o-m+\kappa N|\leq N-1}}^{\infty} \frac{1}{m^{1-D}}\frac{1}{o^{1-D}}\frac{1}{|o-m|^{1-2\epsilon}}.$$

By setting $p := o - m + \kappa N$ the claim follows with (7.3). □

Acknowledgements

This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt A1, C1) of the German Research Foundation (DFG). The authors are also grateful to three unknown referees and the associate editor. Their constructive comments on two earlier versions of this paper resulted in a substantial improvement of our work. Finally we would like to thank Tobias Kley and Kevin Kokot for computational assistance, Lutz Dümbgen for some helpful discussions about Section 3 and Wilfredo Palma for some constructive discussion on the subject and for making the R-Code in Palma et al. (2013) available to us.

References

Andrews, D. W. K. (1999). Estimation when a parameter is on a boundary. *Econometrica*, 67(6):1341–1383. MR1720781

Arteche, J. and Robinson, P. M. (2000). Semiparametric inference in seasonal and cyclical long memory processes. *Journal of Time Series Analysis*, 21(1):1–25. MR1766171

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

Beran, J. (2009). On parameter estimation for locally stationary long-memory processes. *Journal of Statistical Planning and Inference*, 139:900–915.

Beran, J., Sherman, R., Taqqu, M. S., and Willinger, W. (2006). Long-range dependence in variable-bit-rate video traffic. *IEEE Transactions on Communications*, 43:1566–1579.

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

Bhansali, R. J., Giraitis, L., and Kokoszka, P. S. (2006). Estimation of the memory parameter by fitting fractionally differenced autoregressive models. *Journal of Multivariate Analysis*, 97:2101–2130. MR2301628

Breidt, J., Crato, N., and De Lima, P. (1998). The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, 83:325–348. MR1613867

Chen, Y., Härdle, W., and Pigorsch, U. (2010). Localized realized volatility modeling. *Journal of the American Statistical Association*, 105(492):1376–1393.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *Annals of Mathematical Statistics*, 25(3):573–578. MR0065087

Dahlhaus, R. (1988). Empirical spectral processes and their applications to time series analysis. *Stochastic Process and their Applications*, 30:69–83. MR0968166

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics*, 25(1):1–37.

Dahlhaus, R. and Giraitis, L. (1998). On the optimal segment length for parameter estimates for locally stationary time series. *Journal of Time Series Analysis*, 19(6):629–655.

Dahlhaus, R. and Polonik, W. (2009). Empirical spectral processes for locally stationary time series. *Bernoulli*, 15:1–39.

Deo, R. and Chen, W. (2000). On the integral of the squared periodogram. *Stochastic Process and their Applications*, 85:159–176.

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

Fox, R. and Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series. *The Annals of Statistics*, 14:517–532.

Fryzlewicz, P., Sapatinas, T., and Subba Rao, S. (2006). A Haar-Fisz technique for locally stationary volatility estimation. *Biometrika*, 93:687–704.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4(4):221–238. MR0738585

Gijbels, I., Lambert, A., and Qiu, P. (2007). Jump-preserving regression and smoothing using local linear fitting: a compromise. *Annals of the Institute of*

Statistical Mathematics, 59(2):235–272. MR2394168

Grublyte, I. and Surgailis, D. (2014). Projective stochastic equations and nonlinear long memory. *Advances in Applied Probability*, 46(4):1084–1105.

Haslett, J. and Raftery, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind power ressource. *Applied Statistics*, 38.

Heyde, C. C. and Dai, W. (1996). On the robustness to small trends of estimation based on the smoothed periodogram. *Journal of Time Series Analysis*, 17:141–150. MR1381169

Hidalgo, J. and Soulier, P. (2004). Estimation of the location and exponent of the spectral singularity of a long memory process. *Journal of Time Series Analysis*, 25(1):55–81. MR2042111

Kokoszka, P. S. and Taqqu, M. S. (1995). Fractional ARIMA with stable innovations. *Stochastic Processes and Their Applications*, 60:19–47. MR1362317

Kreiß, J.-P., Paparoditis, E., and Politis, D. N. (2011). On the range of the validity of the autoregressive sieve bootstrap. *Annals of Statistics*, 39.

Künsch, H. (1986). Discrimination between monotonic trends and long-range dependence. *J. Appl. Prob.*, 23:1025–1030. MR0867199

Künsch, H. R. (1987). Statistical aspects of self-similar processes. In Prohorov, Y. and Sazanov, V. V., editors, *Proceedings 1st World Congress of the Bernoulli Society*, volume 1, pages 67–74. VNU Science Press, Utrecht.

Mikosch, T. and Starica, C. (2004). Non-stationarities in financial time series, the long range dependence and the IGARCH effects. *The Review of Economics and Statistics*, 86:378–390.

Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica*, 59(4):1161–1167.

Palma, W. (2007). *Long-Memory Time Series: Theory and Methods.* Wiley, Hoboken, NJ.

Palma, W. and Olea, R. (2010). An efficient estimator for locally stationary Gaussian long-memory processes. *Annals of Statistics.*, 38(5):2958–2997. MR2722461

Palma, W., Olea, R., and Guillermo, F. (2013). Estimation and forecasting of locally stationary processes. *Journal of Forecasting.*, 32(1):86–96.

Perron, P. and Qu, Z. (2010). Long-memory and level shifts in the volatility of stock market return indices. *Journal of Business and Economic Statistics*, 23(2):275–290.

Phillips, P. C. B. and Shimotsu, K. (2004). Local Whittle estimation in nonstationary und unit root cases. *Annals of Statistics*, 32(2):656–692.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

Preuß, P. and Vetter, M. (2013). On discriminating between long-range dependence and non stationarity. *Electronic Journal of Statistics*, 7:2241–2297. MR3108814

Reisen, V. A., Rodrigues, A. L., and Palma, W. (2006). Estimation of seasonal fractionally integrated processes. *Computational Statistics & Data Analysis*, 50(2):568–582. MR2201879

Robinson, P. M. (1994). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics*, 23(5):1630–1661.

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

Roueff, F. and von Sachs, R. (2011). Locally stationary long memory estimation. *Stochastic Processes and their Applications*, 121:813–844. MR2770908

Sen, K., Preuß, P., and Dette, H. (2016). Measuring stationariy in long-memory processes. *Statistica Sinica*, 26:313–357.

Sibbertsen, P. and Kruse, R. (2009). Testing for a change in persistence under long-range dependencies. *Journal of Time Series Analysis*, 30:263–285. MR2512240

Starica, C. and Granger, C. (2005). Nonstationarities in stock returns. *The Review of Economics and Statistics*, 87:503–522.

Velasco, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis*, 20:87–127.

Wu, W. and Zhou, Z. (2014). Simultaneous quantile inference for non-stationary long memory time series. Preprint Univerity of Toronto.

Xia, Z. and Qiu, P. (2012). Jump information criterion for statistical inference in estimating discontinuous curves. *Biometrika*, 99(1):1–12. MR3371012

Yau, C. Y. and Davis, R. A. (2012). Likelihood inference for discriminating between long-memory and change-point models. *Journal of Time Series Analysis*, 33(4):649–664. MR2944844

Yong, C. H. (1974). *Asymptotic Behaviour of Trigonometric Series*. Chinese University, Hong Kong.