Detecting long-range dependence in non-stationary time series

Philip Preuß, Kemal Sen, Holger Dette

Detecting long-range dependence in non-stationary time series

Philip Preuß, Kemal Sen, Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum

Germany

email: philip.preuss@ruhr-uni-bochum.de

email: kemal.sen@ruhr-uni-bochum.de

email: holger.dette@ruhr-uni-bochum.de

December 17, 2013

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 nonparametric 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. This test is working in the spectral domain and uses a sieve of approximating tvFARIMA models to estimate the time varying long-range dependence parameter nonparametrically. 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 nonparametric 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.

1 Introduction

Many time series [like asset volatility or regional temperatures] exhibit a slow decay in the auto correlation 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
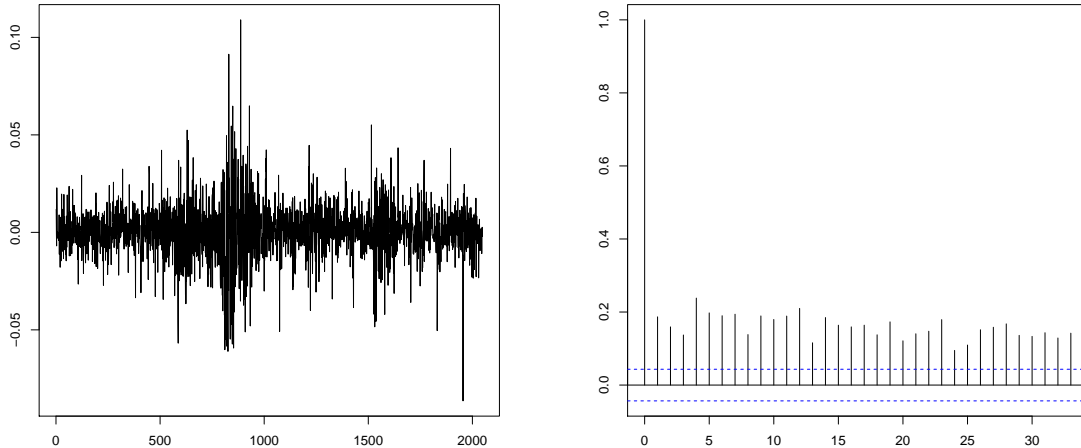
Figure 1: *Left panel: log-returns of the IBM stock between July 15th 2005 and August 30th 2013; right panel: Autocovariance function of the squared returns X_t^2*

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

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

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 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 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 some suitable function, 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) and Sibbertsen and Kruse (2009) developed methods for distinguishing between long-memory and small trends. Furthermore, Berkes et al. (2006), Baek and Pipiras
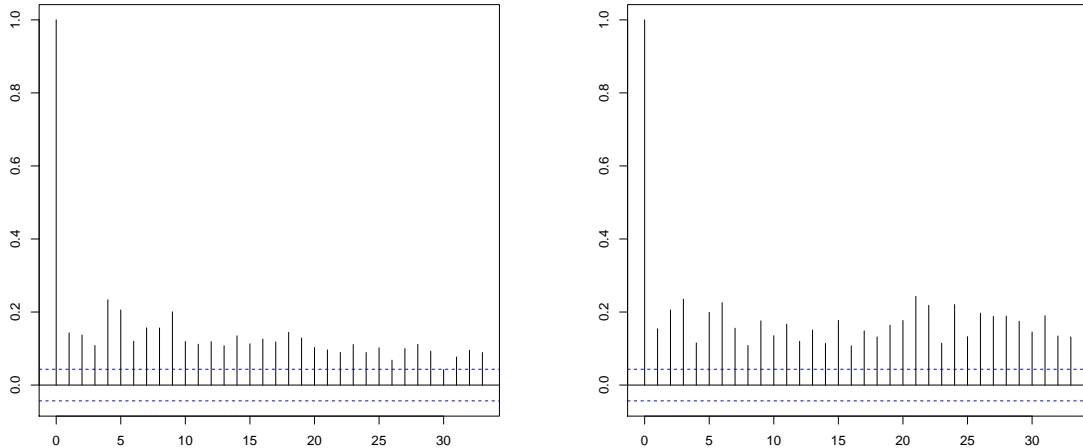
2

Figure 2: *Left panel: 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: 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.*

(2012) and Yau and Davis (2012) investigated CUSUM and likelihood ratio tests to 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 nonparametric 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 is based on an estimate of a (possibly time varying) long-range dependence parameter, which is derived by a sieve of approximating tvFARIMA model. This estimator vanishes if and only if the null hypothesis of a short-memory locally stationary process is satisfied. Its asymptotic properties are investigated in Section 3. In particular we prove asymptotic normality of the proposed 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 4, 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 the application of the method in two data examples. Finally, all technical details are deferred to an appendix.

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 $(\{X_{t,T}\}_{t=1,\dots,T})_{T\in\mathbb{N}}$ of locally stationary long-memory processes, which have an $\mathrm{MA}(\infty)$ representation of the form

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

where

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

$\mu : [0,1] \to \mathbb{R}$ is a "smooth" function and $\{Z_t\}_{t\in\mathbb{Z}}$ are independent standard normal distributed random variables. 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 $(\{X_{t,T}\}_{t=1,\dots,T})_{T\in\mathbb{N}}$ denote a sequence of stochastic processes which have an $\mathrm{MA}(\infty)$ 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

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

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

are satisfied uniformly with respect to u 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 $a : [0,1] \to \mathbb{R}$, $d_0 : [0,1] \to \mathbb{R}_+^0$ in (2.4) are twice continuously differentiable.

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

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

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

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

is twice continuously differentiable.

4

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

$$\sup_{u \in (0,1)} |\psi_l'(u)| \leq C \log |l| |l|^{D-1}, \quad \sup_{u \in (0,1)} |\psi_l''(u)| \leq C \log^2 |l| |l|^{D-1}, \tag{2.8}$$

$$\sup_{u \in (0,1)} \left| \frac{\partial}{\partial u} f(u, \lambda) \right| \leq C \log(\lambda) \lambda^{2D}, \quad \sup_{u \in (0,1)} \left| \frac{\partial^2}{\partial u^2} f(u, \lambda) \right| \leq C \log^2(\lambda) \lambda^{2D}$$

are satisfied.

Similar locally stationary long-range dependent models have been investigated by Beran (2009), Palma and Olea (2010) and Roueff and von Sachs (2011). Note that 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 and allows, for example, to include tvFARIMA(p, d, q)-models as well [see Theorem 2.2 in Preuß and Vetter (2013)]. Moreover, it is also worthwhile to mention that the assumption of Gaussianity is only imposed to simplify the technical arguments in the proofs of our main results and that it is straightforward (but cumbersome) to extend the theory to a more general framework, see Remark 3.9 for more details on this. 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

$$b_0(u) \equiv 1, \quad b_k(u) = \frac{\Gamma(k + d(u))}{\Gamma(d(u))\Gamma(k+1)} \quad \text{and} \quad \left(1 + \sum_{k=1}^{\infty} a_{k,0}(u)z^k\right)^{-1} = \sum_{k=0}^{\infty} a_{k,0}^{(-1)}(u)z^k$$

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].

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

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

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

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

and

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

hold [see Palma and Olea (2010) for details]. Consequently, the autocovariance function $\gamma(u, k) = \text{Cov}(X_0(u), X_k(u))$ is 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.

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 $X_{t,T}$. Since the function d_0 is continuous and non-negative we obtain that the hypotheses

$$\text{H}_0 : d_0(u) = 0 \quad \forall u \in [0, 1] \quad \text{vs.} \quad \text{H}_1 : d_0(u) > 0 \quad \text{for some } u \in [0, 1]$$

are equivalent to

$$\text{H}_0 : F = 0 \quad \text{vs.} \quad \text{H}_1 : F > 0, \tag{2.9}$$

where the quantity F is defined by

$$F := \int_0^1 d_0(u) du. \tag{2.10}$$

In the next section we will develop an estimator of the function d_0 and establish uniform convergency. The integral is then estimated by a Riemann sum and we investigate the asymptotic properties of the resulting estimator both under the null hypothesis and the alternative. In particular we prove consistency and asymptotic normality. As a consequence we obtain a consistent level-α test for the presence of long-range dependence in non-stationary time series models by rejecting the null hypothesis for large values of the estimator of F.

3 Testing short- versus long-memory

In order to estimate the integral F we use a sieve 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

$$f_{\theta_k(u)}(\lambda) = |1 - \exp(i\lambda)|^{-2d(u)} g_k(u, \lambda), \tag{3.1}$$

where

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

and, for each $k \in \mathbb{N}$, $\theta_k = (d, a_1, \ldots, a_k) : [0, 1] \to \mathbb{R}^{k+1}$ is a vector valued function. We then estimate the function $\theta_k(u)$ by a localized Whittle-estimator, that is

$$\hat{\theta}_{N,k}(u) = \arg \min_{\theta_k \in \Theta_{\lfloor uT \rfloor / T, k}} \mathcal{L}_{N,k}^{\hat{\mu}}(\theta_k, u), \tag{3.2}$$

where

$$\mathcal{L}_{N,k}^{\hat{\mu}}(\theta_k, u) := \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\log(f_{\theta_k}(\lambda)) + \frac{I_N^{\hat{\mu}}(u, \lambda)}{f_{\theta_k}(\lambda)} \right) d\lambda \tag{3.3}$$

6

denotes the (local) Whittle likelihood [see Dahlhaus and Giraitis (1998) or Dahlhaus and Polonik (2009)] and for each $u \in [0,1]$ $\Theta_{u,k} \subset \mathbb{R}^{k+1}$ is a compact set which will be specified in Assumption 3.1. In (3.2) and (3.3) the quantity

$$I_N^{\hat{\mu}}(u, \lambda) := \left| \frac{1}{\sqrt{2\pi N}} \sum_{p=0}^{N-1} \left[X_{\lfloor uT \rfloor - N/2 + 1 + p, T} - \hat{\mu}(\lfloor uT \rfloor - N/2 + 1 + p, T) \right] e^{-ip\lambda} \right|^2, \tag{3.4}$$

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 (3.2). It will be demonstrated in Theorem 3.3 below that 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 3.1. *For each $u \in [0,1]$ and $k \in \{k(T), T \in \mathbb{N}\}$ define $\Theta_{u,k} = [0,D] \times \Theta_{u,k,1} \times \ldots \times \Theta_{u,k,k}$, where the constant D is the same as in Assumption 2.1 and 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 $u \in [0,1]$. We assume that the following conditions hold for each $k \in \{k(T), T \in \mathbb{N}\}$:*

(i) The functions g_k in (3.1) 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

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

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

$$\Gamma_k(\theta_k) = \frac{1}{4\pi} \int_{-\pi}^{\pi} f_{\theta_k}^2(\lambda) \nabla f_{\theta_k}^{-1}(\lambda) \nabla f_{\theta_k}^{-1}(\lambda)^T d\lambda, \tag{3.5}$$

$$V_k(\theta_k, u) = \frac{1}{4\pi} \int_{-\pi}^{\pi} f^2(u, \lambda) \nabla f_{\theta_k}^{-1}(\lambda) \nabla f_{\theta_k}^{-1}(\lambda)^T d\lambda,$$

[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

$$\lim_{T \to \infty} \int_0^1 [\Gamma_k^{-1}(\theta_{0,k}(u))]_{1,1} du \Big/ \int_0^1 [\Gamma_k^{-1}(\theta_{0,k}(u)) V_k(\theta_{0,k}(u), u) \Gamma_k^{-1}(\theta_{0,k}(u))]_{1,1} du = 1 \tag{3.6}$$

7

as $T \to \infty$. Furthermore, (3.6) 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

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

is satisfied as $T \to \infty$.

(iv) Define $\Theta_{R,k} = \bigcup_{u \in [0,1]} \Theta_{u,k}$, then

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

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 in order to be fully nonparametric it is necessary that the number of parameters k grows with increasing sample size. In this case the restriction on the spectral norm in part (iv) was verified for a large number of long-range dependent models by Bhansalia et al. (2006) [see equation (4.4) in this reference]. While these assumptions solely depend on the "true" underlying model, the following assumption links the growth rate of k and N as the sample size T increases if the spectral density $f(u, \lambda)$ in (2.6) is approximated by the truncated analogue

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

Assumption 3.2. *Suppose that $N \to \infty$, $N \log(N) = o(T)$ and*

$$\sup_{u \in [0,1]} \sum_{j=k+1}^\infty |a_{j,0}(u)| = O(N^{-1+\varepsilon}) \tag{3.7}$$

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

It follows by similar arguments as given in the proof of Lemma 2.4 in Kreiß et al. (2011) that condition (3.7) implies

$$\sup_{u \in [0,1]} \int_{-\pi}^\pi |f(u, \lambda) - f_{\theta_{0,k}(u)}(\lambda)| d\lambda = O(N^{-1+\varepsilon}). \tag{3.8}$$

As a consequence Assumption 3.1 (iii) is rather intuitive, because the parametric model (3.1) 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 (3.7) 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 Bhansalia 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 by sieve estimation is uniformly consistent for the (time varying) long-range dependence parameter of the locally stationary process.

Theorem 3.3. *Let Assumption 2.1, 3.1 and 3.2 be satisfied and suppose that the estimator of the mean function* μ *satisfies*

$$N^\varepsilon k^3 \max_{t=1,\dots,T} \left|\mu(t/T) - \hat{\mu}(t/T)\right| = o_p(1) \tag{3.9}$$

for some $\varepsilon > 0$. *If* $N^{5/2}/T^2 \to 0$ *and* $k^4 \log^2(T) N^{-\varepsilon/2} \to 0$, *then*

$$\sup_{u\in[0,1]} \left\|\hat{\theta}_{N,k}(u) - \theta_{0,k}(u)\right\|_2 = O_P(k^{3/2} N^{-1/2+\varepsilon} + N^\varepsilon k^{3/2} \max_{t=1,\dots,T} \left|\mu(t/T) - \hat{\mu}(t/T)\right|). \tag{3.10}$$

In particular

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

Remark 3.4. It follows from the proof of Theorem 3.7 below that, the local window estimator with window length N

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

satisfies

$$N^{1/2-D-\alpha} \max_{t=1,\dots,T} \left|\mu(t/T) - \hat{\mu}(t/T)\right| = o_p(1) \quad \text{for every } \alpha > 0.$$

Because we use tvFARIMA$(k, d, 0)$ models in (3.1), we can choose a logarithmic rate for the dimension k [see the discussion following (3.8)]. Consequently, for the local window estimate the uniform rate in equation (3.10) is arbitrary close to the factor $N^{D-1/2}$.

In order to obtain an estimator of the quantity F in (2.10) 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 $t_j = (N-1)j + N/2$, $u_j = t_j/T$, 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

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

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

Theorem 3.5. *Assume that the null hypothesis* H_0 *(of no long-range dependence) is true. Let Assumptions 2.1, 3.1 and 3.2 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*

$$\max_{t=1,\dots,T} \left|\mu(t/T) - \hat{\mu}(t/T)\right| = O_p(N^{-1/2+\varepsilon/2}), \tag{3.11}$$

$$\max_{t=1,\dots,T} \left|\left\{\mu\left(\frac{t-1}{T}\right) - \hat{\mu}\left(\frac{t-1}{T}\right)\right\} - \left\{\mu\left(\frac{t}{T}\right) - \hat{\mu}\left(\frac{t}{T}\right)\right\}\right| = O_p(N^{-1/2-2\varepsilon}T^{-1/2}), \tag{3.12}$$

9

where $0 < \varepsilon < 1/6$ is the constant in Assumption 3.2. Moreover, if the conditions

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

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

$$\sqrt{T}\hat{F}_T/\sqrt{W_T} \xrightarrow{D} \mathcal{N}(0,1). \tag{3.13}$$

Theorem 3.6. *Assume that the alternative H_1 of long-range dependence is true. Let Assumptions 2.1, 3.1 and 3.2 be satisfied and suppose that the estimator $\hat{\mu}$ of the mean function satisfies*

$$N^\varepsilon k^3 \max_{t=1,\dots,T} \left| \mu(t/T) - \hat{\mu}(t/T) \right| = o_p(1), \tag{3.14}$$

where $0 < \varepsilon < \min\{1/2 - D, 1/6\}$ is the constant in Assumption 3.2. Moreover, if the conditions

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

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

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

Note that the term W_T in the denominator of the left hand side in (3.13) can be consistently estimated by

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

such that $\hat{W}_T/W_T \xrightarrow{P} 1$. Consequently, an asymptotic level α-test is obtained by rejecting the null hypothesis (2.9), whenever

$$\sqrt{T}\hat{F}_T/\sqrt{\hat{W}_T} \geq u_{1-\alpha}, \tag{3.15}$$

where $u_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the standard normal distribution. It then follows from Theorem 3.5 and 3.6 that for any estimator of the mean function μ satisfying (3.11), (3.12) and (3.14), the test, which rejects H_0 whenever (3.15) is satisfied, is a consistent level-α test for the null hypothesis stated in (2.9). A popular estimate for this quantity is given by the the local-window estimator

$$\hat{\mu}_L(u) = \frac{1}{L} \sum_{p=0}^{L-1} X_{\lfloor uT \rfloor - L/2 + 1 + p, T} \tag{3.16}$$

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 3.5 and 3.6 if L grows at a 'slightly' faster rate than N. This means, it can be used in the asymptotic level α test defined by (3.15)

Theorem 3.7.

   a) *Suppose that the assumptions of Theorem 3.5 hold and additionally $N^{1+4\varepsilon}/L^{1-\delta} \to 0$ and $L^{5/2-\delta}/T^{3/2} \to 0$ are satisfied for some $\delta > 0$, where $\varepsilon > 0$ denotes the constant in Theorem 3.5. Then the local-window estimator $\hat{\mu}_L$ defined in (3.16) satisfies (3.11) and (3.12).*

   b) *Suppose the assumptions of Theorem 3.6 hold. If additionally $N^{\varepsilon}k^5/L^{1/2-D-\delta} \to 0$ and $L^{5/2-D}/T^2 \to 0$ for some $0 < \delta < 1/2 - D - \varepsilon$ (with the constant ε from Theorem 3.6), then the local-window estimator $\hat{\mu}_L$ defined in (3.16) satisfies (3.14).*

Remark 3.8. Analogues of Theorem 3.5 and 3.6 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 (3.1), 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 3.1(iii) and 3.2 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 3.5 and 3.6 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 3.9. It is worthwhile to mention that the assumption of Gaussianity for the innovation process in 2.1 is not necessary at all and is only imposed here to simplify technical arguments in the proof of Theorem 5.1. In fact, all results of this section remain true as long as the innovations are independent with all moments existing, mean zero and $\mathbb{E}(Z_{t,T}^2) = \sigma^2(t/T)$ 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 5.1 (which is one of the main ingredients for the proofs in Section 5) 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 (5.5) and $\kappa_4(u)$ denotes the fourth cumulant of the innovations, i.e. $\kappa(t/T) = \mathbb{E}(Z_{t,T}^4) - 3\sigma^4(t/T)$ for all t, T. Since we apply this result with $\phi_T(u_j, \lambda) = (4\pi)^{-1}[\Gamma_k^{-1}(\theta_{0,k}(u_j))\nabla f_{\theta_{0,k}(u_j)}^{-1}(\lambda)]_{11}$ in the proof of Theorem 3.5, we must replace $\hat{W}_T$ in the decision rule (3.15) by

$$\hat{W}_{T,general} = \hat{W}_T + \frac{1}{4\pi M} \sum_{j=1}^{M} \hat{\kappa}_4(u_j)/\hat{\sigma}^4(u_j) \Big(\int_{-\pi}^{\pi} f_{\hat{\theta}_{N,k}(u_j)}(\lambda)[\Gamma_k^{-1}(\hat{\theta}_{N,k}(u_j))\nabla f_{\hat{\theta}_{N,k}(u_j)}^{-1}(\lambda)]_{11}d\lambda \Big)^2,$$

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

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

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

11

4 Finite sample properties

The application of the test (3.15) 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 (3.16)] we use $L = N^{1.05}$. Because the procedure is based on a sieve 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

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

where $\lambda_j = 2\pi j/T$ $(j = 1, \ldots, T)$, 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

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

Finally, the performance of the test depends on the choice of N, and this dependency will be investigated in the following discussion.

4.1 Simulation results

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 (3.15) considering three examples. The first example is given by a tvAR(1)-process

$$X_{t,T} = \mu_i(t/T) + 0.6\frac{t}{T}X_{t-1,T} + Z_{t,T}, \quad t = 1, \ldots, T, \tag{4.1}$$

where $\{Z_{t,T}\}_{t=1,\ldots,T}$ denotes a Gaussian white noise process with variance 1. Two cases are investigated for the mean function representing a smooth change and abrupt change in the mean effect, i.e.

$$\mu_1(t/T) = 1.2\frac{t}{T}, \tag{4.2}$$

$$\mu_2(t/T) = \begin{cases} 0.65 & \text{for } t = 1, \ldots, T/2 \\ 1.3 & \text{for } t = T/2 + 1, \ldots T. \end{cases} \tag{4.3}$$

Our third example consists of a tvMA(1)-process given by

$$X_{t,T} = Z_{t,T} + 0.55 \sin\left(\pi\frac{t}{T}\right) Z_{t-1,T}, \quad t = 1, \ldots, T, \tag{4.4}$$

where $\{Z_{t,T}\}_{t=1,\ldots,T}$ denotes again a Gaussian white noise process with variance 1. Figure 3 shows the autocovariance functions of 1024 observations generated by the models (4.2), (4.3) and (4.4), respectively, from which it is clearly visible that the mean functions in (4.2) and (4.3) are causing a long-memory type behaviour of the autocovariance functions [see the left and middle panel in Figure 3]. In Table 1, we show the simulated level of the test (3.15) for various choices of N. We observe a reasonable approximation of the nominal level whenever $N/T \approx 1/4$ and the sample size T is larger or equal than 512. Note that even for model (4.1) with mean
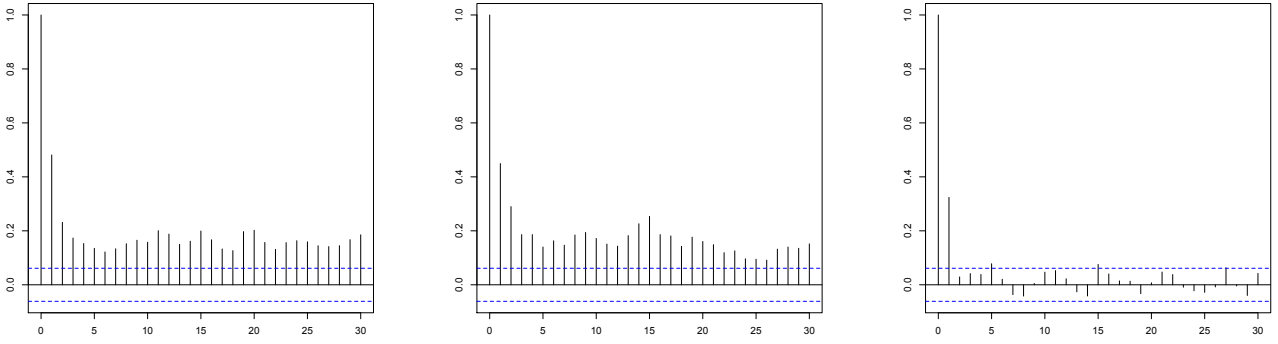
12

Figure 3: *Estimated autocovariance functions of model (4.1) with mean function (4.2) (left panel), (4.3) (middle panel) and of model (4.4) (right panel). The sample size is T=1024.*

function (4.3) the level is only slightly overestimated, although this mean function is not twice continuously differentiable as required by the asymptotic theory. We conjecture that the performance of the test could be improved by using estimators addressing for such jumps.

In order to investigate the power of the test (3.15) 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

$$(1 + 0.2\frac{t}{T}B)(1 - B)^{d(t/T)}X_{t,T} = Z_{t,T}, \quad t = 1, \ldots, T, \tag{4.5}$$

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

$$(1 - B)^{d(t/T)}X_{t,T} = (1 - 0.35\frac{t}{T}B)Z_{t,T}, \quad t = 1, \ldots, T, \tag{4.6}$$

where $\{Z_{t,T}\}_{t=1,\ldots,T}$ denotes a Gaussian white noise process with variance 1 and B is the backshift operator given by $B^j X_{t,T} := X_{t-j,T}$. 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

$$(1 + 0.25B)(1 - B)^{0.1}X_T = (1 - 0.3B)Z_{t,T}, \quad t = 1, \ldots, T. \tag{4.7}$$

The corresponding results for the new test (3.15) are shown in the second column of Table 2, 3 and 4, and we observe reasonable rejection frequencies in the first two cases. 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 2 and 3). The results in Table 4 show a loss in power, which corresponds to intuition because the "average" long-memory effect in model (4.7) is 0.1, while it is $\int_0^1 (0.1 + 0.3u)\, du = 0.25$ in model (4.5) and (4.6).

In order to compare our 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 has the long memory property". The third column of Table 2, 3 and

13

4 shows 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 (3.15) 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 model (4.5)-(4.7). 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 used $q(n) = 15 \log_{10}(n)$ as suggested by these authors in Section 3 of their article. The results are depicted in the fourth column of Table 2, 3 and 4. From this we see that their 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.

Finally, we investigate the method proposed by Yau and Davis (2012) which 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 are depicted in the fifth column of Table 2, 3 and 4 corresponding to non-stationary and stationary long-range dependent alternatives, respectively. We observe that in these cases the new test (3.15) 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 2 and 3). On the other hand, in the case of the long-range dependent stationary alternative (4.7) the test of Yau and Davis (2012) yields slightly better rejection probabilities than the new test (3.15) 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 nonparametric test still yields an improvement.

4.2 Data examples

As an illustration we apply our test to two different datasets. The first contains annual flow volume of the Nile River at Aswan Egypt between the years 1871 and 1970 while the second data set contains 2048 squared log-returns of the IBM stock between July 15th 2005 and August 30th 2013 which were already discussed in the introduction. Both time series are depicted in Figure 4, and in the case of Nile River data our test statistic $\sqrt{T}\hat{F}_T/\sqrt{\hat{W}_T}$ equals -1.9 for $M = 4$, which is far below every reasonable critical value and yields a p-value of 0.971. This implies that the null hypothesis of a non-stationary short-memory model can not be rejected for this dataset, which is in line with the findings of Yau and Davis (2012) who obtained p-values larger than 0.7 for their likelihood ratio approach and the CUSUM procedure of Berkes et al. (2006). The test of Baek and Pipiras (2012) does not reject the null hypothesis as well, since the p-value equals 0.944.

			(4.1), (4.2)		(4.1),(4.3)		(4.4)	
T	N	M	5%	10%	5%	10%	5%	10%
256	64	4	0.090	0.128	0.094	0.145	0.085	0.122
256	32	8	0.151	0.228	0.165	0.255	0.182	0.261
512	128	4	0.061	0.095	0.070	0.114	0.069	0.099
512	64	8	0.089	0.130	0.089	0.126	0.081	0.107
1024	256	4	0.046	0.072	0.077	0.119	0.069	0.106
1024	128	8	0.059	0.087	0.061	0.088	0.064	0.093

Table 1: *Simulated level of the test* (3.15) *for different processes and choices of T,N and M.*

			(3.15)		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						

Table 2: *Rejection frequencies of the test* (3.15) *and three competing procedures under the alternative (4.5) for different choices of T,N and M.*

			(3.15)		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						

Table 3: *Rejection frequencies of the test* (3.15) *and three competing procedures under the alternative (4.6) for different choices of T,N and M.*

			(3.15)		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						

Table 4: *Rejection frequencies of the test* (3.15) *and three competing procedures under the alternative (4.7) for different choices of T,N and M.*
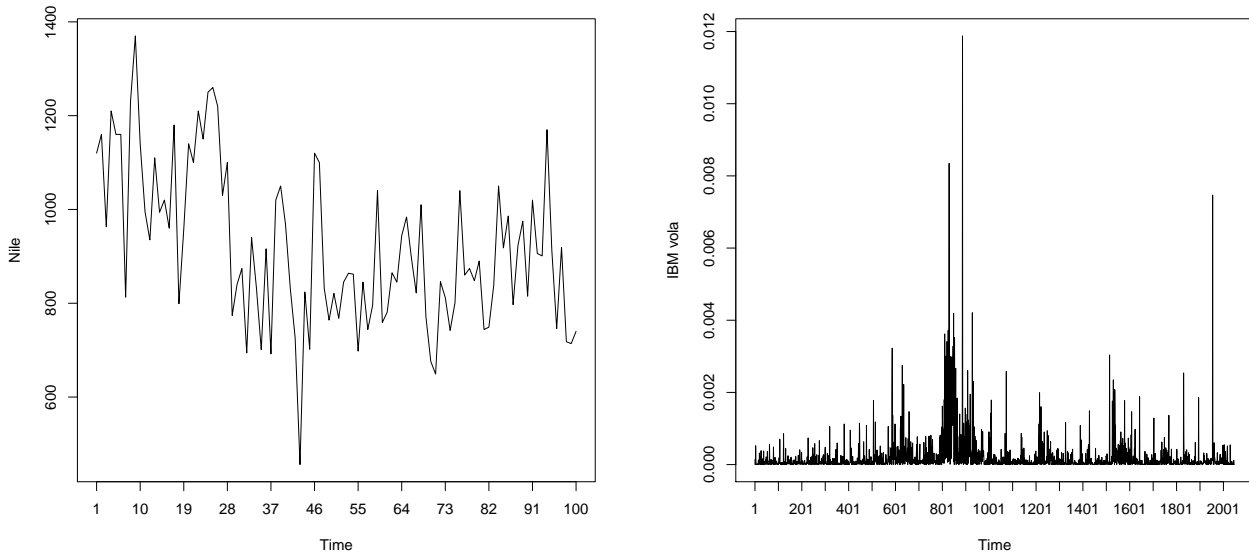
Figure 4: *Left panel: plot of the 100 Nil River Annual flow volume between 1871 and 1970; Right panel: plot of the squared log-returns of the IBM stock between July 15th 2005 and August 30th 2013.*

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 3.9. 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 of the standard normal distribution. This yields an even smaller p-value. 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).

References

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.

16

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

Bhansalia, 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.

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.

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

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

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.

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.

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.

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

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.

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.

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.

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.

Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, 23:1630–61.

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

Sen, K., Preuß, P., and Dette, H. (2013). Measuring stationarity in long-memory processes. Technical report, arXiv:1303.3482, Ruhr-University Bochum.

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

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

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.

5 Appendix: Proofs

5.1 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

$$G_T(\phi_T) = \frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} f(u_j, \lambda) \phi_T(u_j, \lambda) \, d\lambda,$$

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

where I_N^μ is the analogue of the local periodogram (3.4) where the estimator $\hat{\mu}$ has been replaced by the "true" mean function μ.

Theorem 5.1.

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 derivates such that for all $u \in [0,1]$, $\lambda \in [-\pi, \pi]$, $T \in \mathbb{N}$

$$\phi_T(u, \lambda) \leq Cg(k)|\lambda|^{2d_0(u)-\varepsilon}, \tag{5.1}$$

$$\frac{\partial}{\partial \lambda} \phi_T(u, \lambda) \leq Cg(k)|\lambda|^{2d_0(u)-1-\varepsilon}, \tag{5.2}$$

$$\frac{\partial^2}{\partial \lambda^2} \phi_T(u, \lambda) \leq Cg(k)|\lambda|^{2d_0(u)-2-\varepsilon}, \tag{5.3}$$

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

$$\mathbb{E}[\hat{G}_T(\phi_T)] = G_T(\phi_T) + O\left(\frac{g(k)}{N^{1-\varepsilon}}\right) + O\left(\frac{g(k)N^2}{T^2}\right), \tag{5.4}$$

$$\mathrm{Var}[\hat{G}_T(\phi_T)] = V_T + O\left(\frac{1}{T} \frac{g^2(k)}{N^{1-2D-2\varepsilon}}\right) + O\left(\frac{g^2(k)N^2}{T^3}\right) \tag{5.5}$$

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 $D = 0$, $\varepsilon < 1/6$ and additionally $\liminf_{T \to \infty} V_T \geq c$,

$$N \to \infty, \quad g(k)\sqrt{T}/N^{1-\varepsilon} \to 0, \quad g(k)\log(T)/T^{1/6-\varepsilon} \to 0, \quad and \quad g(k)N^2/T^{\frac{3}{2}} \to 0.$$

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

Proof: In order to prove part a) Theorem 5.1 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

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

$$= E_{N,T} + A_{N,T} + B_{N,T}$$

19

where

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

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

$$B_{N,T} := \frac{1}{2\pi N}\frac{1}{M}\sum_{j=1}^{M}\sum_{p,q=0}^{N-1}\sum_{l,m=0}^{\infty}\int_{-\pi}^{\pi}\phi_T(u_j,\lambda)e^{-i(p-q)\lambda}\,d\lambda\mathbb{E}(Z_{t_j,p-l}Z_{t_j,q-m})$$
$$\left\{\big(\psi_{\tilde{t}_j+p,T,l}-\tilde{\psi}_l(u_{j,p})\big)\psi_{\tilde{t}_j+q,T,m}+\tilde{\psi}_l(u_{j,p})\big(\psi_{\tilde{t}_j+q,T,m}-\tilde{\psi}_m(u_{j,q})\big)\right\}.$$

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 (5.4), we prove the following statements:

$$E_{N,T} = \frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}f(u_j,\lambda)\phi_T(u_j,\lambda)\,d\lambda + O\Big(\frac{g(k)}{N^{1-\varepsilon}}\Big) \tag{5.6}$$

$$A_{N,T} = O\Big(\frac{g(k)\log(N)}{N^{1-\varepsilon}M}\Big)+O\Big(\frac{g(k)N^2}{T^2}\Big) \tag{5.7}$$

$$B_{N,T} = O\Big(\frac{g(k)\log(T)}{T}\Big). \tag{5.8}$$

Proof of (5.6): 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 $p \in \{0,1,2\ldots,N-1\}$) which in turn implies $|l-m| \le N-1$. Therefore

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

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

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

where

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

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

Using (2.4), (5.1) and Lemma 6.2 in the online supplement, we obtain

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

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 6.1a) in the online supplement that

$$
\frac{g(k)}{M} \sum_{j=1}^{M} \sum_{\substack{h \in \mathbb{Z} \\ N \leq |h|}} \sum_{\substack{m=1 \\ h+m \geq 1}}^{\infty} \frac{1}{(h+m)^{1-d_0(u_j)}} \frac{1}{m^{1-d_0(u_j)}} \frac{1}{|h|^{1+2d_0(u_j)-\varepsilon}} \leq C g(k) \sum_{\substack{h \in \mathbb{Z} \\ N \leq |h|}} \frac{1}{|h|^{2-\varepsilon}} = O\left(\frac{g(k)}{N^{1-\varepsilon}}\right).
$$

By proceeding analogously we obtain that $E_{N,T}^2 = O(g(k)N^{-1+\varepsilon})$ which proves the assertion in (5.6).

Proof of (5.7): Without loss of generality we only consider the first summand

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

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 (5.6) yield

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

where

$$
A_{N,T}^1 = \frac{1}{2\pi N} \frac{1}{M} \sum_{j=1}^{M} \sum_{\substack{l,m=0 \\ |l-m| \leq N-1}}^{\infty} \sum_{\substack{q=0 \\ 0 \leq q+l-m \leq N-1}}^{N-1} \psi_l(u_j)\psi_m'(u_j)\left(\frac{-N/2+1+q}{T}\right) \int_{-\pi}^{\pi} \phi_T(u_j,\lambda)e^{-i(l-m)\lambda} \, d\lambda,
$$

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

and $\eta_{m,j,q} \in (u_j - N/(2T), u_j + N/(2T))$. Using (2.4), (2.8), (5.1), Lemma 6.2 it follows

$$
|A_{N,T}^1| \leq C \frac{g(k)}{N} \frac{1}{M} \sum_{j=1}^{M} \sum_{\substack{l,m=1 \\ 1 \leq |l-m| \leq N-1}}^{\infty} \frac{1}{l^{1-d_0(u_j)}} \frac{\log(m)}{m^{1-d_0(u_j)}} \frac{1}{|l-m|^{1+2d_0(u_j)-\varepsilon}} \left| \sum_{\substack{q=0 \\ 0 \leq q+l-m \leq N-1}}^{N-1} \left(\frac{-N/2+1+q}{T}\right) \right|
$$

$$
\leq C \frac{g(k)}{T} \frac{1}{M} \sum_{j=1}^{M} \sum_{\substack{l,m=1 \\ 1 \leq |l-m| \leq N-1}}^{\infty} \frac{1}{l^{1-d_0(u_j)}} \frac{\log(m)}{m^{1-d_0(u_j)}} \frac{1}{|l-m|^{2d_0(u_j)-\varepsilon}}
$$

$$
= C \frac{g(k)}{T} \frac{1}{M} \sum_{j=1}^{M} \sum_{\substack{s \in \mathbb{Z} \\ 1 \leq |s| \leq N-1}} \sum_{\substack{l=1 \\ 1 \leq l-s}}^{\infty} \frac{1}{l^{1-d_0(u_j)}} \frac{\log(l-s)}{(l-s)^{1-d_0(u_j)}} \frac{1}{|s|^{2d_0(u_j)-\varepsilon}}
$$

$$\leq C\frac{g(k)\log(N)}{T}\frac{1}{M}\sum_{j=1}^{M}\sum_{\substack{s\in\mathbb{Z}\\1\leq|s|\leq N-1}}\frac{1}{|s|^{1-\varepsilon}}=O\Big(\frac{g(k)\log(N)}{N^{1-\varepsilon}M}\Big)$$

where we used Lemma 6.1(c) in the online supplement for the last step. Finally (2.4), (2.8), (5.1), Lemma 6.2 in the online supplement and the same same arguments as above, show that the term $A_{N,T}^2$ is of order $O(g(k)N^2T^{-2})$.

Proof of (5.8): 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 (5.5). We obtain

$$\mathrm{Var}(\hat{G}_T(\phi_T)) = \frac{1}{(2\pi N)^2}\frac{1}{M^2}\sum_{j_1,j_2=1}^{M}\sum_{p,q,r,s=0}^{N-1}\sum_{l,m,n,o=0}^{\infty}\psi_l(u_{j_1})\psi_m(u_{j_1})\psi_n(u_{j_2})\psi_o(u_{j_2})$$

$$\mathrm{cum}(Z_{t_{j_1},p-l}Z_{t_{j_1},q-m},Z_{t_{j_2},r-n}Z_{t_{j_2},s-o})\int_{-\pi}^{\pi}\phi_T(u_{j_1},\lambda_1)e^{-i(p-q)\lambda_1}\,d\lambda_1\int_{-\pi}^{\pi}\phi_T(u_{j_2},\lambda_2)e^{-i(r-s)\lambda_2}\,d\lambda_2$$

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

where we used assumption (2.3) and similar arguments as given in the proof of (5.4). Because of the Gaussianity of the innovations we obtain

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

$$+\mathbb{E}(Z_{t_{j_1},p-l}Z_{t_{j_2},s-o})\mathbb{E}(Z_{t_{j_1},q-m}Z_{t_{j_2},r-n}).$$

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

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

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 $|l-n+(j_2-j_1)N|\leq N-1$ and $|o-m+(j_1-j_2)N|\leq N-1$. Thus, the term $V_{N,T}^1$ can be written as

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

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

Since $q \in \{0, 1, 2 \ldots, N-1\}$, we get from the condition $0 \leq q + o - m + (j_1 - j_2)N \leq N-1$ 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 6.3 (i)–(iii) in the online supplement that there appears an error of order $O(\frac{1}{T} \frac{g^2(k)}{N^{1-2D-2\varepsilon}})$ if we drop the condition $0 \leq r + l - n + (j_2 - j_1)N \leq N-1$ and assume that the variable r runs from $-(N-1)$ to -1. Therefore, up to an error of order $O(\frac{1}{T} \frac{g^2(k)}{N^{1-2D-2\varepsilon}})$, the term $V^1_{N,T}$ is equal to

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

where

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

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

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

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

We show

$$
D_{1,T} = \frac{2\pi}{N} \frac{1}{M^2} \sum_{j_1=1}^{M} \int_{-\pi}^{\pi} f^2(u_{j_1}, \lambda_1)\phi_T^2(u_{j_1}, \lambda_1) \, d\lambda_1 + O\Big(\frac{1}{T} \frac{g^2(k)}{N^{1-2D-2\varepsilon}}\Big) \tag{5.9}
$$

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

which then concludes the proof of (5.5). 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 $0 \leq q + o - m + (j_1 - j_2)N \leq N-1$.

II. We can drop the condition $|l - n + (j_2 - j_1)N| \leq N-1$ by making an error of order $O(g^2(k)T^{-1}N^{-1+2D+2\varepsilon})$ [this follows from Lemma 6.3(iv) in the online supplement].

III. There appears an error of order $O(g^2(k)T^{-1}N^{-1+2D+2\varepsilon})$ if we omit the sum with $j_1 \neq j_2$ [we prove this in Lemma 6.3(v) in the online supplement].

IV. We can afterwards omit the condition $0 \leq q + o - m \leq N-1$ since it is $0 \leq r \leq N-1$ 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 $|o - m| \leq N-1$ since $r - q + m - o = 0$ and $|r - q| \leq N-1$.

23

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 $O(g^2(k)T^{-1}N^{-1+2D+2\varepsilon}))$

$$\frac{1}{N^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{q,r=0}^{N-1}\int_{-\pi}^{\pi}f(u_{j_1},\lambda_1)\phi_T^2(u_{j_1},\lambda_1)e^{-i(r-q)\lambda_1}\,d\lambda_1\int_{-\pi}^{\pi}f(u_{j_1},\lambda_2)e^{-i(r-q)\lambda_2}\,d\lambda_2$$

$$=\frac{1}{N^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{s=-(N-1)}^{N-1}\int_{-\pi}^{\pi}f(u_{j_1},\lambda_1)\phi_T^2(u_{j_1},\lambda_1)e^{-is\lambda_1}\,d\lambda_1\int_{-\pi}^{\pi}f(u_{j_1},\lambda_2)e^{-is\lambda_2}\,d\lambda_2(N-|s|)$$

$$=D_{1,T}^{(1)}+D_{1,T}^{(2)}+D_{1,T}^{(3)},$$

where

$$D_{1,T}^{(1)}=\frac{1}{N}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{s=-\infty}^{\infty}\int_{-\pi}^{\pi}f(u_{j_1},\lambda_1)\phi_T^2(u_{j_1},\lambda_1)e^{-is\lambda_1}\,d\lambda_1\int_{-\pi}^{\pi}f(u_{j_1},\lambda_2)e^{-is\lambda_2}\,d\lambda_2$$

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

$$D_{1,T}^{(3)}=-\frac{1}{N^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{s=-(N-1)}^{N-1}|s|\int_{-\pi}^{\pi}f(u_{j_1},\lambda_1)\phi_T^2(u_{j_1},\lambda_1)e^{-is\lambda_1}\,d\lambda_1\int_{-\pi}^{\pi}f(u_{j_1},\lambda_2)e^{-is\lambda_2}\,d\lambda_2$$

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 6.2 in the online supplement yields (up to a constant) the inequalities

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

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

which proves (5.9). We now consider the term

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

where

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

24

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

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

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

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 $O(T^{-1}g^2(k)N^{-1+2D+2\varepsilon})$ because of Lemma 6.3 (vi) in the online supplement.

Part b) follows with par a) if we show

$$\text{cum}_l[\sqrt{T}\hat{G}_T(\phi)] = O\big(g(k)^l T^{l(\varepsilon-1/2+2D)+(1-4D)} \log(T)^l\big) \quad \text{for } l \geq 3 \text{ and } D < 1/4. \tag{5.10}$$

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

$$\frac{1}{N} \sum_{k=1}^{N/2} \phi_T(u, \lambda_k) e^{ih\lambda_k} = O\Big(\frac{1}{|h \text{ modulo } N/2|}\Big) \tag{5.11}$$

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 (5.10) with the only difference that $\varepsilon = 0$ and $g(k) \equiv 1$. In our situation, assumption (5.1) and Lemma 6.2 in the online supplement imply

$$\int_{-\pi}^{\pi} \phi_T(u, \lambda) e^{ih\lambda} d\lambda = O\Big(\frac{g(k)}{|h|^{1+2d_0(u)-\varepsilon}}\Big) = O\Big(T^\varepsilon \frac{g(k)}{|h|}\Big) \tag{5.12}$$

and we can therefore proceed completely analogously to the proof of Theorem 6.1 c) in Preuß and Vetter (2013) but using (5.12) instead of (5.11). The details are omitted for the sake of brevity. □

For the formulation of the next result we define the set

$$\mathcal{G}_T(s, \ell) = \{\tilde{\phi}_T : [-\pi, \pi] \to \mathbb{R} \mid \tilde{\phi}_T \text{ is symmetric, there exists a polynomial } P_\ell \text{ of degree } \ell \text{ and a}$$
$$\text{constant } d \in [0, 1/2) \text{ such that } \tilde{\phi}_T(\lambda) = \log^s(|1 - e^{i\lambda}|)|1 - e^{i\lambda}|^{2d}|P_\ell(e^{i\lambda})|^2\}$$

and state the following result.

Theorem 5.2. *Suppose Assumption 2.1 and 3.2 are fulfilled, $N^{5/2}/T^2 \to 0$ and $0 < \varepsilon < 1/2 - D$ is the constant of Assumption 3.2. 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 (5.1)–(5.3) 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 $k^4 \log^2(T)N^{-\varepsilon/2} \to 0$. Then*

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

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 5.2 is a consequence of the statements:

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

$$\hat{G}_{T,1}(\phi_T^*) := N^{1/2-\varepsilon/2}\Big(\hat{D}_{T,1}(\phi_T^*) - \int_{-\pi}^{\pi} f(t/T,\lambda)\phi_T(t/T,\lambda)d\lambda\Big) = o_p(1) \tag{5.13}$$

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

$$\lim_{T\to\infty} P\Big(\sup_{\phi_{T,1}^*,\phi_{T,2}^* \in \Phi_T^*} |\hat{G}_{T,1}(\phi_{T,1}^*) - \hat{G}_{T,1}(\phi_{T,2}^*)| > \eta\Big) = 0. \tag{5.14}$$

In order to prove part (i) we use the same arguments as given in the proof of (5.4) and (5.5) and obtain

$$\mathbb{E}[\hat{D}_{T,1}(\phi_T^*))] = \int_{-\pi}^{\pi} f(t/T,\lambda)\phi_T(t/T,\lambda)d\lambda + O\Big(\frac{1}{N^{1-\varepsilon}}\Big) + O\Big(\frac{N^2}{T^2}\Big),$$

$$\mathrm{Var}[N^{1/2}\hat{D}_{T,1}(\phi_T^*)] = \int_{-\pi}^{\pi} f^2(t/T,\lambda)\phi_T^2(t/T,\lambda)d\lambda + O\Big(\frac{1}{N^{1-2D-2\varepsilon}}\Big) + O\Big(\frac{N^2}{T^2}\Big),$$

which yields (5.13) 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 inequlality

$$\mathbb{E}(|\hat{G}_{T,1}(\phi_{T,1}^*) - \hat{G}_{T,1}(\phi_{T,2}^*)|^l) \le (2l)!C^l \Delta_{T,\varepsilon}^l(\phi_{T,1}^*,\phi_{T,2}^*)$$

holds for all even $l \in \mathbb{N}$ and all $\phi_{T,1}^*,\phi_{T,2}^* \in \Phi_T^*$, where

$$\Delta_{T,\varepsilon}(\phi_{T,1}^*,\phi_{T,2}^*) = 1_{\{t_1=t_2\}}N^{-\varepsilon/2}\sqrt{\int_{-\pi}^{\pi} f^2(t_1/T,\lambda)(\phi_{T,1,1}(t_1/T,\lambda) - \phi_{T,1,2}(t_1/T,\lambda))^2 d\lambda} + A1_{\{t_1\neq t_2\}}N^{-\varepsilon/2}$$

for a constant A which is sufficiently large such that

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

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

$$P(|\hat{G}_{T,1}(\phi_{T,1}^*) - \hat{G}_{T,1}(\phi_{T,2}^*)| > \eta) \le 96\exp(-\sqrt{\eta\Delta_{T,\varepsilon}^{-1}(\phi_{T,1}^*,\phi_{T,2}^*)C^{-1}})$$

for all $\phi_{T,1}^*,\phi_{T,2}^* \in \Phi_T^*$. The statement (5.14) 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 $\Delta_{T,\varepsilon}$ is finite. More precisely, the covering number $N_T(u)$ of Φ_T^* with respect to $\Delta_{T,\varepsilon}$ is equal to one for $u \ge AN^{-\varepsilon/2}$ and bounded by $TC^{(qk)^2}u^{-qk}N^{-qk\varepsilon/2}$ for some constant C for $u < AN^{-\varepsilon/2}$ [see Chapter VII.2. of Pollard (1984) for a definition of covering numbers]. This implies that the covering integral $J_T(\delta) = \int_0^\delta \log(48N_T(u)^2u^{-1})^2 du$ is up to a constant bounded by $k^4\log^2(T)N^{-\varepsilon/2}$. The assertion follows by the assumptions on k and N. $\qquad\square$

5.2 Proof of Theorem 3.3

Introducing the notation

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

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

$$\max_{t=1,\ldots,T} \left| \mathcal{L}_{N,k}^{\hat{\mu}}(\theta_k, t/T) - \mathcal{L}_{N,k}^{\mu}(\theta_k, t/T) \right|$$

$$\leq C \max_{t=1,\ldots,T} \max_{q=0,\ldots,N} \left\{ \left| \mu(t/T) - \hat{\mu}(t/T) \right| \left| \int_{-\pi}^{\pi} d_N^{X-\mu}(t/T, \lambda) f_{\theta_k}^{-1}(\lambda) e^{iq\lambda} d\lambda \right| \right\} + C \max_{t=1,\ldots,T} \left| \mu(t/T) - \hat{\mu}(t/T) \right|^2$$

for some constant $C \in \mathbb{R}$ and $d_N^{X-\mu}$ is defined by $\left| d_N^{X-\mu}(u, \lambda) \right|^2 := I_N^{\mu}(u, \lambda)$. By proceeding as in the proof of Theorem 5.2 one verifies

$$\max_{t=1,\ldots,T} \max_{q=0,\ldots,N} \sup_{\theta_k \in \Theta_{R,k}} \left| \int_{-\pi}^{\pi} d_N^{X-\mu}(t/T, \lambda) f_{\theta_k}^{-1}(\lambda) e^{iq\lambda} d\lambda \right| = O(N^{\varepsilon}),$$

and (3.9) yields

$$\max_{t=1,\ldots,T} \sup_{\theta_k \in \Theta_{R,k}} \left| \mathcal{L}_{N,k}^{\hat{\mu}}(\theta_k, t/T) - \mathcal{L}_{N,k}^{\mu}(\theta_k, t/T) \right| = \max_{t=1,\ldots,T} \left| \mu(t/T) - \hat{\mu}(t/T) \right| O_p(N^{\varepsilon}) = o_p(k^{-5/2}), \quad (5.15)$$

and analogously we get

$$\max_{t=1,\ldots,T} \sup_{\theta_k \in \Theta_{R,k}} \left\| \nabla \mathcal{L}_{N,k}^{\hat{\mu}}(\theta_k, t/T) - \nabla \mathcal{L}_{N,k}^{\mu}(\theta_k, t/T) \right\|_2 = \max_{t=1,\ldots,T} \left| \mu(t/T) - \hat{\mu}(t/T) \right| O_p(k^{1/2} N^{\varepsilon}) = o_p(k^{-5/2}). (5.16)$$

For each $u \in [0, 1]$ let $\hat{\theta}_{N,k}(u)$ denote the Whittle-estimator defined in (3.2). Then Theorem 5.2 and similar arguments as in the proof of Theorem 3.2 in Dahlhaus (1997) yield

$$\sup_{u \in [0,1]} \left\| \hat{\theta}_{N,k}(u) - \theta_{0,k}(u) \right\|_2 = o_p(1). \tag{5.17}$$

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

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

and the first term on the left-hand side vanishes due to (5.17). This yields

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

where E_T denotes the difference between $\nabla \mathcal{L}_{N,k}^{\mu}(\theta_{0,k}(u), u)$ and $\nabla \mathcal{L}_{N,k}^{\hat{\mu}}(\theta_{0,k}(u), u)$, which is of order $\max_{t=1,\ldots,T} \left| \mu(t/T) - \hat{\mu}(t/T) \right| O_p(k^{1/2} N^{\varepsilon})$ by (5.16). It follows from

$$\nabla \mathcal{L}_{N,k}^{\mu}(\theta_k, u) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[I_N^{\mu}(u, \lambda) - f_{\theta_k}(\lambda) \right] \nabla f_{\theta_k}^{-1}(\lambda) d\lambda$$

and Theorem 5.2 that $\max_{u \in \{1/T, \ldots, 1\}} \|\nabla \mathcal{L}_{N,k}^{\mu}(\theta_{0,k}(u), u)\|_2 = O_p(\sqrt{k} N^{-1/2+\varepsilon/2})$ so it remains to show

$$P(\nabla^2 \mathcal{L}_{N,k}^{\hat{\mu}}(\zeta_u^{(k)}, u)^{-1} \text{ exists and } \|\nabla^2 \mathcal{L}_{N,k}^{\hat{\mu}}(\zeta_u^{(k)}, u)^{-1}\|_{sp} \leq Ck \text{ for all } u \in \{1/T, \ldots, 1\}) \to 1$$

for some positive constant C. This, however, follows with a Taylor expansion, (5.17), Theorem 5.2 and Assumption 3.1 (iv) for the corresponding expression with $\hat{\mu}$ replaced by μ. The more general case is then implied by the convergence-assumptions on $\hat{\mu}$. $\qquad \square$

5.3 Proof of Theorem 3.5 and Theorem 3.6

We will show in Section 5.3.1 that under the null hypothesis H_0 the estimate

$$\max_{j=1,\ldots,M} \left\|\hat{\theta}_{N,k}(u_j) - \theta_{0,k}(u_j)\right\|_2 = O_p(k^{3/2}N^{-1/2+\varepsilon/2}) \tag{5.18}$$

is valid, while Theorem 3.3 and (3.14) imply

$$k^{3/2} \max_{j=1,\ldots,M} \left\|\hat{\theta}_{N,k}(u_j) - \theta_{0,k}(u_j)\right\|_2 = o_p(1) \tag{5.19}$$

under the alternative H_1. As in the proof of Theorem 3.3 there exist vectors $\zeta_j^{(k)} = (\zeta_{j,1}^{(k)}, \zeta_{j,2}^{(k)}, \ldots, \zeta_{j,k+1}^{(k)}) \in \mathbb{R}^{k+1}$, $j = 1, \ldots, M$, satisfying $\|\zeta_j^{(k)} - \theta_{0,k}(u_j)\|_2 \leq \|\hat{\theta}_{N,k}(u_j) - \theta_{0,k}(u_j)\|_2$ such that

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

holds because of Assumption 3.1 (ii) and (5.18) (under H_0) or (5.19) (under H_1). By rearranging and summing over every block, it follows that

$$\frac{1}{M}\sum_{j=1}^{M} (\hat{\theta}_{N,k}(u_j) - \theta_{0,k}(u_j)) = R_{0,T} - R_{1,T} - R_{2,T} - R_{3,T} - R_{4,T} \tag{5.20}$$

where

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

Γ_k^{-1} is defined in (3.5) and the terms $R_{i,T}(i = 1\ldots,4)$ are given by

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

$$R_{2,T} := \frac{1}{M}\sum_{j=1}^{M} \Gamma_k^{-1}(\theta_{0,k}(u_j))(\nabla^2\mathcal{L}_{N,k}^{\hat{\mu}}(\zeta_j^{(k)}, u_j) - \nabla^2\mathcal{L}_{N,k}^{\mu}(\zeta_j^{(k)}, u_j))(\hat{\theta}_{N,k}(u_j) - \theta_{0,k}(u_j)),$$

$$R_{3,T} := \frac{1}{M}\sum_{j=1}^{M} \Gamma_k^{-1}(\theta_{0,k}(u_j))(\nabla^2\mathcal{L}_{N,k}^{\mu}(\zeta_j^{(k)}, u_j) - \nabla^2\mathcal{L}_{N,k}^{\mu}(\theta_{0,k}(u_j), u_j))(\hat{\theta}_{N,k}(u_j) - \theta_{0,k}(u_j)),$$

$$R_{4,T} := \frac{1}{M}\sum_{j=1}^{M} \Gamma_k^{-1}(\theta_{0,k}(u_j))(\nabla^2\mathcal{L}_{N,k}^{\mu}(\theta_0(u_j), u_j) - \Gamma_k(\theta_{0,k}(u_j)))(\hat{\theta}_{N,k}(u_j) - \theta_{0,k}(u_j)).$$

We obtain for the first summand in (5.20)

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

and with the notation $\phi_T(u_j,\lambda) = 1/(4\pi)[\Gamma_k^{-1}(\theta_{0,k}(u_j))\nabla f_{\theta_{0,k}(u_j)}^{-1}(\lambda)]_{11}$, it is easy to see that Assumption 3.1 (i)–(iv) imply the conditions of Theorem 5.1 b) with $g(k) = k^2$. Moreover, observing the definition of V_T and

28

W_T in Theorem 5.1 and 3.5, (3.6) yields $V_T/W_T \to 1$. Consequently, under the assumptions of Theorem 3.5 it follows (observing (3.8) and the growth conditions on N, T)

$$\frac{\sqrt{T}}{M} \sum_{j=1}^{M} \left[\Gamma_k^{-1}(\theta_{0,k}(u_j)) \nabla \mathcal{L}_N^\mu(\theta_{0,k}(u_j), u_j) \right]_{11} / \sqrt{W_T} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Since $d_0(u)$ is the first element of the vector $\theta_{0,k}(u)$, Theorem 3.5 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

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

Analogously, Theorem 3.6 follows from (5.4) and (5.5) if the estimates

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

can be established. It can be shown analogously to the proof of Theorem 3.6 in Dahlhaus (1997), that, under assumptions (3.11) – (3.12), both terms $R_{1,T}$ and $R_{2,T}$ are of order $O_p(k^2 N^{-\varepsilon}T^{-1/2} + k^2 N^{\varepsilon-1})$, while, under assumption (3.14), the order is $o_p(1)$ [see the proof of (5.23) and (5.15), respectively, for more details]. Therefore it only remains to consider the quantities $R_{3,T}$ and $R_{4,T}$. For this purpose note that

$$\nabla^2 \mathcal{L}_{N,k}^\mu(\theta_k(u_j), u_j) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[I_N^\mu(u_j, \lambda) - f_{\theta_k(u_j)}(\lambda) \right] \nabla^2 f_{\theta_k(u_j)}^{-1}(\lambda) \, d\lambda + \Gamma_k(\theta_k(u_j)) \tag{5.21}$$

$$\nabla^3 \mathcal{L}_{N,k}^\mu(\theta_k(u_j), u_j) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[I_N^\mu(u_j, \lambda) - f_{\theta_k(u_j)}(\lambda) \right] \left[\frac{\partial^3 f_{\theta_k(u_j)}^{-1}(\lambda)}{\partial \theta_{j,t} \partial \theta_{j,s} \partial \theta_{j,r}} \right]_{r,s,t=1,\dots,k+1} d\lambda$$

$$- \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[\frac{\partial f_{\theta_k(u_j)}(\lambda)}{\partial \theta_{j,t}} \frac{\partial^2 f_{\theta_k(u_j)}^{-1}(\lambda)}{\partial \theta_{j,s} \partial \theta_{j,r}} \right]_{r,s,t=1,\dots,k+1} d\lambda$$

$$+ \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[\frac{\partial}{\partial \theta_{j,t}} \left(\frac{\partial f_{\theta_k(u_j)}(\lambda)}{\partial \theta_{j,s}} \frac{1}{f_{\theta_k(u_j)}^2(\lambda)} \frac{\partial f_{\theta_k(u_j)}(\lambda)}{\partial \theta_{j,r}} \right) \right]_{r,s,t=1,\dots,k+1} d\lambda, \tag{5.22}$$

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 $\|Ax\|_2 \le \|A\|_{sp}\|x\|_2$

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

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

$$\|\nabla^2 \mathcal{L}_{N,k}^\mu(\zeta_j^k, u_j) - \nabla^2 \mathcal{L}_{N,k}^\mu(\theta_0(u_j), u_j)\|_{sp} \le k \max_{r,s=1,\dots,k} \left| \left[\nabla^2 \mathcal{L}_{N,k}^\mu(\zeta_j^{(k)}, u_j) - \nabla^2 \mathcal{L}_{N,k}^\mu(\theta_{0,k}(u_j), u_j) \right]_{r,s} \right|$$

$$= k \max_{r,s=1,\dots,k} \left| \nabla \left[\nabla^2 \mathcal{L}_{N,k}^\mu(\tilde{\zeta}_j^{(k)}, u_j) \right]_{r,s} (\zeta_j^{(k)} - \theta_{0,k}(u_j)) \right| \le k \max_{r,s=1,\dots,k} \left\| \nabla \left[\nabla^2 \mathcal{L}_{N,k}^\mu(\tilde{\zeta}_j^{(k)}, u_j) \right]_{r,s} \right\|_2 \|\zeta_j^{(k)} - \theta_{0,k}(u_j)\|_2$$

$$\le k \|\hat{\theta}_{N,k}(u_j) - \theta_{0,k}(u_j)\|_2 \sup_{\substack{\theta_k \in \Theta_{R,k} \\ r,s=1,\dots,k}} \left\| \nabla \left[\nabla^2 \mathcal{L}_{N,k}^\mu(\theta_k, u_j) \right]_{r,s} \right\|_2,$$

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

$$\|R_{3,T}\|_2 \le k \max_{j=1,\dots,M} \|\hat{\theta}_{N,k}(u_j) - \theta_{0,k}(u_j)\|_2^2 \sup_{\theta_k \in \Theta_{R,k}} \|\Gamma_k^{-1}(\theta_k)\|_{sp} \sup_{\substack{\theta_k \in \Theta_{R,k}; j=1,\dots,M \\ r,s=1,\dots k}} \left\| \nabla \left[\nabla^2 \mathcal{L}_N^\mu(\theta_k, u_j) \right]_{r,s} \right\|_2$$

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

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

where, in the last inequality, we have used the fact that the second and third term in (5.22) are bounded by a constant [this follows directly from Assumption 3.1]. Before we investigate the order of this expression, we derive a similar bound for the term $R_{4,T}$. Observing (5.21) we obtain

$$\|R_{4,T}\|_2 \leq \max_{j=1,\dots,M} \left\| \hat{\theta}_{N,k}(u_j) - \theta_{0,k}(u_j) \right\|_2 \sup_{\theta_k \in \Theta_{R,k}} \left\| \Gamma_k^{-1}(\theta_k) \right\|_{sp} \max_{j=1,\dots,M} \left\| \nabla^2 \mathcal{L}_{N,k}^{\mu}(\theta_{0,k}(u_j), u_j) - \Gamma_k(\theta_{0,k}(u_j)) \right\|_{sp}$$

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

$$\times \max_{j=1,\dots,M} \left\| \frac{1}{4\pi} \int_{-\pi}^{\pi} [I_N^{\mu}(u_j, \lambda) - f_{\theta_{0,k}(u_j)}(\lambda)] \nabla^2 f_{\theta_{0,k}(u_j)}^{-1}(\lambda) \, d\lambda \right\|_{sp}$$

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

$$\times \max_{j=1,\dots,M} \max_{r,s=1,\dots,k} \left| \frac{1}{4\pi} \int_{-\pi}^{\pi} [I_N^{\mu}(u_j, \lambda) - f_{\theta_{0,k}(u_j)}(\lambda)] \frac{\partial^2 f_{\theta_{0,k}(u_j)}^{-1}(\lambda)}{\partial \theta_{j,s} \partial \theta_{j,r}} \, d\lambda \right|.$$

If we show

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

$$\max_{j=1,\dots,M} \max_{r,s=1,\dots,k} \left| \frac{1}{4\pi} \int_{-\pi}^{\pi} [I_N^{\mu}(u_j, \lambda) - f_{\theta_{0,k}(u_j)}(\lambda)] \frac{\partial^2 f_{\theta_{0,k}(u_j)}^{-1}(\lambda)}{\partial \theta_{j,s} \partial \theta_{j,r}} \, d\lambda \right| = O_p(N^{-1/2+\varepsilon/2}),$$

it follows with Assumption 3.1 (iv) in combination with (5.18) (under H_0) and (5.19) (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 5.2 and (3.8). $\square$

5.3.1 Proof of (5.18)

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

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

$$\times \left(\max_{t=1,\dots,T} \left| \left\{ \mu \left(\frac{t-1}{T} \right) - \hat{\mu} \left(\frac{t-1}{T} \right) \right\} - \left\{ \mu \left(\frac{t}{T} \right) - \hat{\mu} \left(\frac{t}{T} \right) \right\} \right| + \max_{t=1,\dots,T} |\mu(t/T) - \hat{\mu}(t/T)|/N \right)$$

$$\Pi_{2,T} = C \sup_{\theta_k \in \Theta_{R,k}} |f_{\theta_k}^{-1}(\lambda)| \max_{t=1,\dots,T} |\mu(t/T) - \hat{\mu}(t/T)|^2$$

and C denotes a positive constant. By proceeding as in the proof of Theorem 5.2 one obtains

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

which implies (observing the assumptions (3.11) and (3.12))

$$\max_{j=1,...,M} \sup_{\theta_k \in \Theta_{R,k}} \left| \mathcal{L}_{N,k}^{\hat{\mu}}(\theta_k, u_j) - \mathcal{L}_{N,k}^{\mu}(\theta_k, u_j) \right| = O_p(N^{-\varepsilon}T^{-1/2} + N^{\varepsilon-1}) = o_P(N^{-1/2+\varepsilon/2}k^{1/2}) \tag{5.23}$$

under H_0. Analogously we obtain

$$\max_{j=1,...,M} \sup_{\theta_k \in \Theta_{R,k}} \left\| \nabla \mathcal{L}_{N,k}^{\hat{\mu}}(\theta_k, u_j) - \nabla \mathcal{L}_{N,k}^{\mu}(\theta_k, u_j) \right\|_2$$

$$= O_p(k^{1/2} N^{-\varepsilon} T^{-1/2} + k^{1/2} N^{\varepsilon-1}) = o_P(N^{-1/2+\varepsilon/2}k^{1/2}) \tag{5.24}$$

under the null hypothesis. By using (5.23) and (5.24) instead of (5.15) and (5.16), assertion (5.18) follows by the same arguments as given in the proof of Theorem 3.3. $\qquad\square$

5.4   Proof of Theorem 3.7

A second order Taylor expansion yields

$$\mathbb{E}(\hat{\mu}_L(t/T)) = \mu(t/T) + \frac{\mu'(t/T)}{L} \sum_{p=0}^{L-1} (-L/2 + 1 + p)/T + O(L^2/T^2) = \mu(t/T) + O(1/T + L^2/T^2). \tag{5.25}$$

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

$$\text{cum}(\hat{\mu}_L(t_1/T), \hat{\mu}_L(t_2/T), ..., \hat{\mu}_L(t_l/T)) = \frac{1}{L^l} \sum_{p_1,...,p_l=0}^{L-1} \sum_{m_1,...,m_l=0}^{\infty} \psi_{t,T,m_1} \cdots \psi_{t,T,m_l} \text{cum}(Z_{p_1-m_1}, ..., Z_{p_l-m_l})$$

are bounded by

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

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 (5.25) and the bound $C^l L^{1-l(1-D)}$ for the l-th cumulant ($l \geq 2$) of $\hat{\mu}_L$ yield $\text{cum}_l(L^{1/2-D-\alpha/2}(\hat{\mu}_L(t_1/T) - \hat{\mu}_L(t_2/T))) \leq (2C)^l L^{-l\alpha/2}$ for all $t_i \in \{1, ..., T\}$ and every $\alpha > 0$, from which we get

$$\mathbb{E}(L^{l(1/2-D-\alpha)}(\hat{\mu}_L(t_1/T) - \hat{\mu}_L(t_2/T))^l) \leq (2l)! C^l L^{-l\alpha/2} \quad \text{for all even } l \in \mathbb{N} \text{ and } t_i \in \{1, ..., T\}$$

[see the proof of Lemma 2.3 in Dahlhaus (1988) for more details]. By considering the order of the bias (5.25) this yields

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

31

as in the proof of Theorem 5.2. Consequently (3.11) [under the conditions of part a)] and (3.14) [under the conditions of part b)] follow. So it remains to show (3.12) in the case $D = 0$. For this purpose we define

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

and from (5.25) 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 $l \geq 3$. This yields, as above,

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

for every $\alpha > 0$, and completes the proof of Theorem 3.7. $\qquad\square$

6   Online supplement: Auxiliary results

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

Lemma 6.1. *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*

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

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

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

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. (2013). $\qquad\square$

Lemma 6.2. *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*

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

uniformly in T.

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

Lemma 6.3. *If Assumption 2.1 holds, then*

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

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

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

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

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

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

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

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

(v)
$$\frac{1}{N^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{r,q=0}^{N-1}\sum_{l,m,n,o=0}^{\infty}\sum_{\substack{j_2=1\\|j_1-j_2|\geq1\\0\leq q+o-m+(j_1-j_2)N\leq N-1\\|o-m+(j_1-j_2)N|\leq N-1}}^{M}\psi_l(u_{j_1})\psi_m(u_{j_1})\psi_n(u_{j_2})\psi_o(u_{j_2})$$

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

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

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

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

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), (5.1) and Lemma 6.2 to bound the term in (i) (up to a constant) through

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

$$\frac{1}{|r-q+l-n+(j_2-j_1)N|^{1+2d_0(u_{j_1})-\varepsilon}}\frac{1}{|r-q+m-o+(j_2-j_1)N|^{1-\varepsilon}}.$$

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 $l^{-1 d_0(u_{j_1})}m^{-1+d_0(u_{j_1})}n^{-1+D}o^{-1+D}$, 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

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

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

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

$$\times\frac{1}{(s-q+m)^{1-D}}\frac{1}{|g-q|^{1+2d_0(u_{j_1})-\varepsilon}}\frac{1}{|r-s|^{1-\varepsilon}}$$

Through an repeated application of (7.1) and (7.2) the claim now follows.

Proof of (v): By setting

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

34

we can write the term in (v) as

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

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

and by integrating over λ_2 this is the same as

$$\frac{4\pi^2}{N^2}\frac{1}{M^2}\sum_{j_1=1}^{M}\sum_{q=0}^{N-1}\sum_{m,o=0}^{\infty}\sum_{\substack{j_2=1\\|j_1-j_2|\geq 1\\0\leq q+o-m+(j_1-j_2)N\leq N-1\\|o-m+(j_1-j_2)N|\leq N-1}}^{M}\psi_m(u_{j_1})\psi_o(u_{j_2})\int_{-\pi}^{\pi}\phi_T(u_{j_1},\lambda_1)\phi_T(u_{j_2},\lambda_1)f(u_{j_1},u_{j_2},\lambda_1)e^{-i(o-m)\lambda_1}\,d\lambda_1.$$

By (5.1) and Lemma 6.2 this sum can be bounded by

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

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

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

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

By setting $p := o - m + \kappa N$ the claim follows with (7.2). $\qquad\square$