Testing semiparametric hypotheses in locally stationary processes

Philip Preuß, Mathias Vetter, Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum

Germany

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

email: mathias.vetter@ruhr-uni-bochum.de

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

February 24, 2011

Abstract

In this paper we investigate the problem of testing semiparametric hypotheses in locally stationary processes. The proposed method is based on an empirical version of the L_2-distance between the true time varying spectral density and its best approximation under the null hypothesis. As this approach only requires estimation of integrals of the time varying spectral density and its square, we do not have to choose a smoothing bandwidth for the local estimation of the spectral density – in contrast to most other procedures discussed in the literature. Asymptotic normality of the test statistic is derived both under the null hypothesis and the alternative. We also propose a bootstrap procedure to obtain critical values in the case of small sample sizes. Additionally, we investigate the finite sample properties of the new method and compare it with the currently available procedures by means of a simulation study. Finally, we illustrate the performance of the new test in a data example investigating log returns of the S&P 500.

1 Introduction

Most of the classical literature on time series analysis assumes that the underlying process is second-order stationary. While this assumption is quite tempting from a mathematical point of view, because it allows for an elegant development of useful statistical inference procedures, it is often difficult to justify in applications, as most processes in reality change their dependency structure over time. A more realistic framework is therefore one that allows for changes in the second-order characteristics over time and many stochastic models which address this issue have been developed. Out of the large literature we mention the early work on this subject of Priestley (1965), who considered oscillating processes. In the last decade the concept of locally stationary processes has become quite popular, because in contrast to other concepts this approach allows for a meaningful asymptotic theory, which is essential for statistical inference in such models. Locally stationary processes were introduced by Dahlhaus (1996) as processes with an $MA(\infty)$ representation, where the coefficients can vary smoothly over time. An important subclass is that of time varying $ARMA(p,q)$ models.

While estimation procedures for locally stationary processes in various settings have found considerable interest in the literature [see for example Neumann and von Sachs (1997), Dahlhaus et al. (1999), Chang and Morettin (1999), Dahlhaus and Polonik (2006) or Van Bellegem and von Sachs (2008) among others], the problem of testing parametric or semiparametric hypotheses [like testing for a $tvARMA(p,q)$ structure] has received much less attention. However, testing hypotheses of this type is very important, because if a model has been misspecified the use of model-based estimation and forecasting procedures may yield an inefficient and – in the worst case – inconsistent and invalid statistical analysis of the data.

In the "classical" case of stationary processes this statistical problem has been studied intensively in the literature [see for example Hjellvik et al. (1998), Paparoditis (2000), Dette and Spreckelsen (2003), Delgado et al. (2005), Eichler (2008) among many others]. For Gaussian locally stationary processes, likelihood ratio as well as Wald and Lagrange multiplier tests of the null hypothesis that the time varying spectral density depends on a finite dimensional parameter vector were suggested in Sakiyama and Taniguchi (2003). As pointed out by Sergides and Paparoditis (2009), this approach is quite restrictive and does not allow to test for a $tvARMA(p,q)$ model for example. These authors proposed an alternative test to decide whether a locally stationary process belongs to a general semiparametric class of time varying processes against an unspecified, locally stationary alternative. Their method is based on an L_2-distance between a nonparametric and a semiparametric estimate of the local spectral density.

A drawback in the approach of Sergides and Paparoditis (2009) is the choice of two regularization parameters. The first one controls the length of the intervals on which the local estimation of the spectral density takes place. This regularization is inherent to any statistical inference in locally stationary processes and cannot be avoided. The second additional regularization parameter occurs in a kernel smoothing procedure which is used for the nonparametric estimation of the local spectral density. As

it was pointed out in Sergides and Paparoditis (2009), it is especially the choice of the smoothing bandwidth in this estimate that influences the properties of the corresponding test substantially.

This paper is devoted to the presentation of a simple alternative method for testing semiparametric hypotheses which does not require the choice of an additional smoothing bandwidth. As in Sergides and Paparoditis (2009) our approach is based on an empirical L_2-distance measure, but it avoids kernel smoothing by estimating integrals of certain functionals of the local spectral density directly instead of estimating the local spectral density first. As a consequence, the procedure proposed in this paper requires only the specification of one regularization parameter, which is impossible to avoid in such kinds of testing problems.

The remaining part of the paper is organized as follows. In Section 2 we introduce the necessary notation and explain the basic principle of the method. In particular, we derive the asymptotic distribution of the proposed statistic both under the null hypothesis and the alternative. In Section 3 we introduce a bootstrap method to approximate the distribution of the test statistic under the null hypothesis when sample sizes are rather small. The finite sample properties of the new test are studied in Section 4 by means of a simulation study, where we also provide a comparison with the procedure proposed by Sergides and Paparoditis (2009) and analyze 2049 log returns of the S&P 500. Finally, all technical details are deferred to an appendix in Section 5.

2 The testing procedure

Following Dahlhaus and Polonik (2009), we define locally stationary time series via a sequence of stochastic processes $\{X_{t,T}\}_{t=1,\ldots,T}$, where each random variable $X_{t,T}$ exhibits a linear representation of the form

$$(2.1) \qquad X_{t,T} = \sum_{l=-\infty}^{\infty} \psi_{t,T,l} Z_{t-l}, \quad t = 1, \ldots, T$$

and the random variables Z_t are i.i.d. $N(0,1)$ distributed [the assumption of Gaussianity is imposed to simplify technical arguments, see Remark 2.9]. Since the constants $\psi_{t,T,l}$ are in general time dependent, each process $\{X_{t,T}\}$ is typically not stationary. However, if we interpret $X_{t,T}$ as the observation of some underlying process X at time t/T and if we further assume that the coefficients $\psi_{t,T,l}$ behave like some smooth functions in the neighbourhood of each t/T, then the time series can be regarded as locally stationary in the sense that observations nearby show approximately stationary behaviour. For this reason we impose additionally that there exist twice continuously differentiable functions $\psi_l : [0,1] \to \mathbb{R}$ $(l \in \mathbb{Z})$ with

$$(2.2) \qquad \sum_{l=-\infty}^{\infty} \sup_{t=1,\ldots,T} |\psi_{t,T,l} - \psi_l(t/T)| = O(1/T).$$

3

Furthermore, we assume that the technical conditions

(2.3)
$$\sum_{l=-\infty}^{\infty} \sup_{u \in [0,1]} |\psi_l(u)||l|^2 < \infty,$$

(2.4)
$$\sum_{l=-\infty}^{\infty} \sup_{u \in [0,1]} |\psi_l'(u)||l| < \infty,$$

(2.5)
$$\sum_{l=-\infty}^{\infty} \sup_{u \in [0,1]} |\psi_l''(u)| < \infty$$

are satisfied, which are in general rather mild, as they hold in a variety of time-varying $\mathrm{ARMA}(p,q)$ models. Note also that variables Z_t with time varying variance $\sigma^2(t/T)$ can be included in the model by choosing other coefficients $\psi_{t,T,l}$ in the representation (2.1) appropriately.

The uniquely defined function

$$f(u, \lambda) = \frac{1}{2\pi} |\psi(u, \exp(-i\lambda))|^2$$

with

$$\psi(u, \exp(-i\lambda)) := \sum_{l=-\infty}^{\infty} \psi_l(u) \exp(-i\lambda l)$$

is called the time varying spectral density of $\{X_{t,T}\}$ [see Dahlhaus (1996)]. It is continuous by assumption, and we denote by $\mathcal{F}_{LS}$ the set of all time varying spectral densities which satisfy (2.3)–(2.5).

The aim of this paper is to develop a test for the hypothesis that the time varying spectral density $f(u, \lambda)$ has a semiparametric structure, so the null hypothesis is given by

(2.6)
$$H_0: \quad f(u, \lambda) \in \mathcal{F}_{SPLS} = \{g(\theta(u), \lambda); \theta(u) = (\theta_1(u), ..., \theta_p(u)) \in \Theta\}$$

for some fixed integer p. Here, Θ denotes an appropriately defined space of p-dimensional functions and g is some fixed mapping which will be specified below. Throughout this paper we will make the following basic assumptions regarding the class $\mathcal{F}_{SPLS}$.

Assumption 2.1 For each $u \in [0,1]$ let Θ_u be a compact subset of $\mathbb{R}$ and set $\Theta_R = \bigcup_{u \in [0,1]} \Theta_u$. For Θ, we assume the following condition to hold:

(i) Θ is the space of all continuously differentiable functions θ with $\theta(u) \in \Theta_u$ for all $u \in [0,1]$.

We assume that the function g is defined on $\Theta_R \times \mathbb{R}$. Furthermore,

(ii) For each $\theta \in \Theta_R$, the function $g(\theta, \lambda)$ is symmetric and 2π-periodic in its second component λ.

(iii) The functions g and $1/g$ are four times continuously differentiable, and any of their partial derivatives up to order four is continuous on $\Theta_R \times [-\pi, \pi]$ and uniformly bounded from below and above.

4

Example 2.1 An example of particular importance which fits in the described setup is the null hypothesis that the underlying process is a tvARMA(p,q) process, i.e. $X_{t,T}$ is generated by the equation

$$X_{t,T} + \sum_{j=1}^{p} a_j(t/T)X_{t-j,T} = \varepsilon_{t,T} + \sum_{j=1}^{q} b_j(t/T)\varepsilon_{t-j,T}$$

with independent random variables $\varepsilon_{t,T} \sim N(0, \sigma^2(t/T))$. To avoid identifiability issues we assume that $a_p(u)$ and $b_q(u)$ are not zero everywhere. In Theorem 2.3 of Dahlhaus (1996) it is shown that a member of this class of tvARMA(p,q) processes indeed defines a locally stationary process, if the functions $a_j(\cdot)$, $b_j(\cdot)$ are both continuous on the interval $[0,1]$ and for all $u \in [0,1]$ the polynomial $\sum_{j=0}^{p} a_j(u)z^j \neq 0$ has no zeros in the disc $\{z : |z| \leq 1 + \delta\}$ for some constant $\delta > 0$ which is independent of u. In this case the time varying spectral density is given by

$$f(u,\lambda) = \frac{\sigma^2(u)}{2\pi} \frac{\left| \sum_{j=0}^{q} b_j(u)\exp(i\lambda j) \right|^2}{\left| \sum_{j=0}^{p} a_j(u)\exp(i\lambda j) \right|^2},$$

and we see that $f(u,\lambda)$ is of the form $g(\theta(u), \lambda)$ with $\theta(u) = (a_1(u), ..., a_p(u), b_1(u), ..., b_q(u), \sigma^2(u))$. Moreover, Assumption 2.1 is satisfied for an appropriate choice of the set Θ.

In order to measure the deviation between the true spectral density and its best approximation under the null hypothesis, we proceed in a similar way as Sergides and Paparoditis (2009). Following Dahlhaus (1996), the asymptotic Kullback-Leibler divergence between two Gaussian locally stationary processes with respective spectral densities $k(u,\lambda)$ and $l(u,\lambda)$ is given by

$$\mathcal{L}_{KL}(k,l) = \frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} \left(\log \frac{l(u,\lambda)}{k(u,\lambda)} + \frac{k(u,\lambda)}{l(u,\lambda)} - 1 \right) d\lambda du.$$

This result motivates the following choice of $\theta_0 : [0,1] \to \Theta_R$ as the best approximation of the local spectral density $f(u,\lambda)$ by a given semiparametric class. We set for each $u \in [0,1]$

$$(2.7) \qquad \theta_0(u) = \text{argmin}_{\theta \in \Theta_u} \mathcal{L}(u,\theta),$$

$$(2.8) \qquad \mathcal{L}(u,\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\log g(\theta,\lambda) + \frac{f(u,\lambda)}{g(\theta,\lambda)} \right) d\lambda.$$

This means that we determine the parameter $\theta_0 : [0,1] \to \mathbb{R}^p$ by minimizing a local version of the Kullback-Leibler divergence $\mathcal{L}(f, g_\theta)$, where we ignore terms not depending on θ. If the null hypothesis holds and Assumption 2.1 is satisfied, then θ_0 indeed corresponds to the "parameter" of the true spectral density. Under the alternative it is neither obvious that $\theta_0(u)$ exists nor that the function θ_0 belongs to Θ. For this reason we introduce

Assumption 2.2 We have $f \in \mathcal{F}_{LS}$ and assume that the set $\mathcal{F}_{SPLS}$ satisfies

5

(i) For each $u \in [0, 1]$ the parameter $\theta_0(u)$ defined in (2.7) exists and is unique.

(ii) For each $u \in [0, 1]$ the parameter $\theta_0(u)$ is an element of the interior of Θ_u.

(iii) For each $u \in [0, 1]$ the matrix $\frac{\partial^2 \mathcal{L}(u,\theta)}{\partial \theta^2}$ is invertible at $(u, \theta_0(u))$.

The last assumption guarantees that the minimizing function θ_0 indeed belongs to Θ, which follows from an application of the implicit function theorem. Our idea for the construction of a test for the hypothesis

$$H_0 : f \in \mathcal{F}_{SPLS} \quad \text{vs.} \quad H_1 : f \in \mathcal{F}_{LS} \backslash \mathcal{F}_{SPLS}$$

is to estimate one of the two closely related L_2-distances

$$(2.9) \qquad D^2 := \frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} \Big(f(u, \lambda) - g(\theta_0(u), \lambda) \Big)^2 d\lambda du,$$

$$(2.10) \qquad R^2 := \frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} \Big(\frac{f(u, \lambda)}{g(\theta_0(u), \lambda)} - 1 \Big)^2 d\lambda du,$$

where the parameter $\theta_0(u)$ has been defined in (2.7). Both measures are justified by the fact that we have $f(u, \lambda) = g(\theta_0(u), \lambda)$ under the null hypothesis in which case $D^2 = R^2 = 0$ holds, whereas each of the two distances is positive, whenever $f(u, \lambda)$ corresponds to the alternative H_1. The quantity D^2 is motivated by mathematical concept of L_2-approximation, while R^2 is considered for two reasons: it is scale invariant by construction, and a Taylor expansion of the function $h(z) = \log(z^{-1}) + z - 1$ gives $h(z) = \frac{1}{2}(z - 1)^2 + O((z - 1)^3)$. Consequently, up to a constant the measure R^2 can be regarded as an approximation of the Kullback-Leibler divergence $\mathcal{L}(f, g_{\theta_0})$.

In order to define reasonable empirical versions of D^2 and R^2, we need at least rough estimators for the local spectral density. Those will be obtained from local periodograms, and we assume (without loss of generality) that the total sample size T can be decomposed as $T = NM$, where N and M are integers and N is even. We define the local periodogram at time u by

$$(2.11) \qquad I_N^X(u, \lambda) := \frac{1}{2\pi N} \Big| \sum_{s=0}^{N-1} X_{\lfloor uT \rfloor - N/2 + 1 + s, T} \exp(-i\lambda s) \Big|^2$$

[see Dahlhaus (1997)], where we have set $X_{j,T} = 0$, if $j \notin \{1, \dots, T\}$. This is the usual periodogram computed from the observations $X_{\lfloor uT \rfloor - N/2 + 1, T}, \dots, X_{\lfloor uT \rfloor + N/2, T}$, and this explains why it serves as a reasonable estimate for $f(u, \lambda)$, if N tends to infinity.

An estimator for $\theta_0(u)$ will now be constructed from an empirical version of the distance $\mathcal{L}(u, \theta)$ defined in (2.8). To this end we set

$$\mathcal{L}_T(u, \theta) := \frac{1}{N} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \Big\{ \log g(\theta, \lambda_k) + \frac{I_N(u, \lambda_k)}{g(\theta, \lambda_k)} \Big\},$$

6

so we use a Riemann sum at the Fourier frequencies $\lambda_k := \frac{2\pi k}{N}$ to approximate the integral with respect to λ and replace the unknown $f(u, \lambda_k)$ by the previously defined local periodogram. In a next step we define $\hat{\theta}_T(u) := \operatorname{argmin}_{\theta \in \Theta_u} \mathcal{L}_T(u, \theta)$, and the following Lemma shows that $\hat{\theta}_T(u)$ is a consistent estimator for $\theta_0(u)$.

Lemma 2.3 *If Assumption 2.1, Assumption 2.2 and*

$$(2.12) \qquad\qquad N \to \infty, \quad M \to \infty, \quad \frac{T^{1/2}}{N} \to 0, \quad \frac{N}{T^{3/4}} \to 0$$

are satisfied, then for all $u \in [0, 1]$ we have

$$\hat{\theta}_T(u) = \theta_0(u) + O_p(\sqrt{1/N})$$

as $T \to \infty$.

It is worth noting that Lemma 2.3 holds both under the null hypothesis H_0 and the alternative H_1. This result suggests to replace the unknown function $\theta_0(u)$ by $\hat{\theta}_T(u)$ to obtain empirical versions of D^2 and R^2, and we focus on the estimation of the measure D^2 defined in (2.9) for a moment. We have a suitable approximation for the integral with respect to the variable λ using the Fourier frequencies from above, whereas for the integral with respect to u we introduce the notation

$$u_j := \frac{t_j}{T} := \frac{N(j-1) + N/2}{T}.$$

In other words, we split the interval $[0, 1]$ into M disjoint intervals, each with N observations and midpoint u_j. Thus the entire procedure depends on the particular choice of the regularization parameter N, which is an intrinsic feature of any method for statistical inference in locally stationary processes, as one needs local estimates for the spectral density [see Dahlhaus (1997), Sakiyama and Taniguchi (2004), Sergides and Paparoditis (2008), Sergides and Paparoditis (2009) or Palma and Olea (2010) among many others]. Especially in the problem of testing semiparametric hypotheses it is impossible to avoid some kind of smoothing, because one has to estimate the function $\theta(u)$ in some way. From the discussion above, it follows that a natural estimator for D^2 is given by

$$\hat{F}_T = \frac{1}{T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \left(I_N^X(u_j, \lambda_k) - g(\hat{\theta}_T(u_j), \lambda_k) \right)^2.$$

Our first result regards the error due to the estimation of $\theta(u)$ and is obtained from Lemma 2.3 and a Taylor expansion.

Theorem 2.4 *Suppose that Assumption 2.1, Assumption 2.2 and (2.12) are satisfied. Then we have*

$$\frac{1}{T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \left(g(\theta_0(u_j), \lambda_k) - g(\hat{\theta}_T(u_j), \lambda_k) \right)^2 = O_P(1/N).$$

It turns out that this approximation error is of small order compared to the rate of convergence. Indeed, the latter one will be of order $\sqrt{1/T}$, and from (2.12) we have $\sqrt{T}/N \to 0$. Therefore, we will focus first on the oracle

$$\hat{F}_T^o = \frac{1}{T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \left(I_N^X(u_j, \lambda_k) - g(\theta_0(u_j), \lambda_k) \right)^2.$$

This statistic is not a consistent estimate of D^2, as it contains an additional bias term. Nevertheless, a bias correction can be carried out easily and a consistent estimator for D^2 is given by

$$\hat{G}_T^o := \frac{1}{T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \left(I_N^X(u_j, \lambda_k) - g(\theta_0(u_j), \lambda_k) \right)^2 - \frac{1}{2T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} I_N^X(u_j, \lambda_k)^2.$$

We have precisely

Lemma 2.5 *If Assumption 2.1, Assumption 2.2 and (2.12) are satisfied, then as $T \to \infty$*

$$\hat{F}_T^o \to D^2 + \frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} f^2(u, \lambda) d\lambda du \quad and \quad \hat{G}_T^o \to D^2$$

in probability.

These results suggest that our final estimator for D^2 will be given by the statistic

(2.13)
$$\hat{G}_T := \frac{1}{T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \left(I_N^X(u_j, \lambda_k) - g(\hat{\theta}_T(u_j), \lambda_k) \right)^2 - \frac{1}{2T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} I_N^X(u_j, \lambda_k)^2.$$

Our main result specifies the asymptotic properties of $\hat{G}_T$ both under the null hypothesis and the alternative.

Theorem 2.6 *If Assumption 2.1, Assumption 2.2 and (2.12) are satisfied, then as $T \to \infty$ we have*

$$\sqrt{T}(\hat{G}_T - D^2) \xrightarrow{D} \mathcal{N}(0, \sigma_1^2),$$

where the asymptotic variance is given by

$$\sigma_1^2 = \frac{5}{4\pi} \int_0^1 \int_{-\pi}^{\pi} f^4(u, \lambda) d\lambda du + \frac{1}{\pi} \int_0^1 \int_{-\pi}^{\pi} g^2(\theta_0(u), \lambda) f^2(u, \lambda) d\lambda du - \frac{2}{\pi} \int_0^1 \int_{-\pi}^{\pi} g(\theta_0(u), \lambda) f^3(u, \lambda) d\lambda du.$$

In a similar manner as Lemma 2.5 it can be shown that the statistic

$$\hat{V}_{1,T} := \frac{5}{24T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} I_N^X(u_j, \lambda_k)^4 + \frac{2}{T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} g^2(\hat{\theta}_T(u_j), \lambda_k) I_N^X(u_j, \lambda_k)^2 - \frac{4}{3T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} g(\hat{\theta}_T(u_j), \lambda_k) I_N^X(u_j, \lambda_k)^3$$

defines a consistent estimator of the asymptotic variance in Theorem 2.6. Under the null hypothesis H_0 the asymptotic variance reduces to

$$\sigma_{1,H_0}^2 = \frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} g^4(\theta_0(u), \lambda) d\lambda du,$$

which can easily be estimated by

(2.14)
$$\hat{V}_{1,H_0,T} := \frac{1}{T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} g^4(\hat{\theta}_T(u_j), \lambda_k).$$

Therefore an asymptotic level α test is obtained by rejecting the null hypothesis if

(2.15)
$$\sqrt{T} \frac{\hat{G}_T}{\sqrt{\hat{V}_{1,H_0,T}}} \geq z_{1-\alpha},$$

where $z_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the standard normal distribution. Due to $\hat{G}_T \to D^2$ in probability, the test is consistent as well.

With the same arguments as above it can be shown that the statistic

$$\hat{H}_T := \frac{1}{T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \left(\frac{I_N(u_j, \lambda_k)}{g(\hat{\theta}_T(u_j), \lambda_k)} - 1 \right)^2 - \frac{1}{2T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \frac{I_N^X(u_j, \lambda_k)^2}{g^2(\hat{\theta}_T(u_j), \lambda_k)}$$

defines a consistent estimator for the distance R^2 defined in (2.10) and the analogue to Theorem 2.6 is given by the following result.

Theorem 2.7 *If Assumption 2.1, Assumption 2.2 and (2.12) are satisfied, then as $T \to \infty$ we have*

$$\sqrt{T}(\hat{H}_T - R^2) \xrightarrow{D} \mathcal{N}(0, \sigma_2^2),$$

where the asymptotic variance is given by

$$\sigma_2^2 = \frac{5}{4\pi} \int_0^1 \int_{-\pi}^{\pi} \frac{f^4(u, \lambda)}{g^4(\theta_0(u), \lambda)} d\lambda du + \frac{1}{\pi} \int_0^1 \int_{-\pi}^{\pi} \frac{f^2(u, \lambda)}{g^2(\theta_0(u), \lambda)} d\lambda du - \frac{2}{\pi} \int_0^1 \int_{-\pi}^{\pi} \frac{f^3(u, \lambda)}{g^3(\theta_0(u), \lambda)} d\lambda du.$$

Under the null hypothesis the asymptotic variance simplifies to

$$\sigma_{2,H_0}^2 = \frac{1}{2},$$

whereas in the general case a consistent estimator for the variance σ_2^2 is given by

(2.16)
$$\hat{V}_{2,T} := \frac{5}{24T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \frac{I_N^X(u_j, \lambda_k)^4}{g^4(\hat{\theta}_T(u_j), \lambda_k)} + \frac{2}{T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \frac{I_N^X(u_j, \lambda_k)^2}{g^2(\hat{\theta}_T(u_j), \lambda_k)} - \frac{4}{3T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \frac{I_N^X(u_j, \lambda_k)^3}{g^3(\hat{\theta}_T(u_j), \lambda_k)}.$$

9

Hence an asymptotic level α test is obtained by rejecting the null hypothesis if

$$(2.17) \qquad\qquad \sqrt{2T}\hat{H}_T \geq z_{1-\alpha}.$$

Due to $\hat{H}_T \to R^2$ in probability this test is consistent as well, and in practice it appears to be advantageous to the previous one, as we do not have to estimate the asymptotic variance in Theorem 2.7. However, in the simulation study in Section 4 we will demonstrate that a bootstrap version of the test (2.15) is more robust with respect to the choice of the regularization parameter N than the test defined by (2.17) and therefore both proposals have their pros and cons.

Remark 2.8
(a) Note that instead of $\hat{\theta}_T(u)$ every other $\sqrt{N}$-consistent estimator for $\theta_0(u)$ could be used without affecting the validity of Theorem 2.6 and Theorem 2.7.
(b) From Theorem 2.6 and 2.7 it follows that

$$\left[0, \hat{G}_T + z_{1-\alpha}\sqrt{\hat{V}_{1,T}/T}\right] \quad \text{and} \quad \left[0, \hat{H}_T + z_{1-\alpha}\sqrt{\hat{V}_{2,T}/T}\right]$$

are asymptotic $(1-\alpha)$ confidence interval for the quantities D^2 and R^2, respectively [here $z_{1-\alpha}$ denotes the $(1-\alpha)$ quantile of the standard normal distribution].
(c) By means of Theorem 2.7 we can construct an asymptotic level α test for the so called *precise hypothesis*

$$(2.18) \qquad\qquad H_0 : R^2 > \varepsilon \quad \text{versus} \quad H_1 : R^2 \leq \varepsilon$$

[see Berger and Delampady (1987)] as well, where $\varepsilon > 0$ denotes a prespecified constant. We reject the hypothesis H_0 whenever

$$\hat{H}_T - \varepsilon < z_\alpha\sqrt{\hat{V}_{2,T}/T},$$

where the estimate $\hat{V}_{2,T}$ is defined in (2.16). The importance of the hypothesis (2.18) stems from the fact that in reality a particular model will never be precisely correct. Note that R^2 defines a measure for the deviation from the null hypothesis. Consequently – if the null hypotheses in (2.18) is rejected – the experimenter decides for an approximately semiparametric model at a controlled type I error.

Remark 2.9
We note that the assumption of Gaussianity is only needed to simplify technical arguments. The results in Theorem 2.6 and 2.7 can be extended to the more general case of independent and identically distributed innovations with existing moments of all order. In this case both Theorem 2.6 and 2.7

remain valid with the different asymptotic variances

$$\sigma_{1,g}^2 = \frac{5}{4\pi} \int_0^1 \int_{-\pi}^{\pi} f^4(u,\lambda) d\lambda du + \frac{1}{\pi} \int_0^1 \int_{-\pi}^{\pi} g^2(\theta_0(u),\lambda) f^2(u,\lambda) d\lambda du - \frac{2}{\pi} \int_0^1 \int_{-\pi}^{\pi} g(\theta_0(u),\lambda) f^3(u,\lambda) d\lambda du$$

$$+ \kappa_4 \Big\{ \frac{1}{4\pi^2} \int_0^1 \Big(\int_{-\pi}^{\pi} f^2(u,\lambda) d\lambda \Big)^2 du + \frac{1}{4\pi^2} \int_0^1 \Big(\int_{-\pi}^{\pi} g(\theta_0(u),\lambda) f(u,\lambda) d\lambda \Big)^2 du$$

$$- \frac{1}{2\pi^2} \int_0^1 \Big(\int_{-\pi}^{\pi} f^2(u,\lambda) d\lambda \Big) \Big(\int_{-\pi}^{\pi} g(\theta_0(u),\lambda) f(u,\lambda) d\lambda \Big) du \Big\}$$

and

$$\sigma_{2,g}^2 = \frac{5}{4\pi} \int_0^1 \int_{-\pi}^{\pi} \frac{f^4(u,\lambda)}{g^4(\theta_0(u),\lambda)} d\lambda du + \frac{1}{\pi} \int_0^1 \int_{-\pi}^{\pi} \frac{f^2(u,\lambda)}{g^2(\theta_0(u),\lambda)} d\lambda du - \frac{2}{\pi} \int_0^1 \int_{-\pi}^{\pi} \frac{f^3(u,\lambda)}{g^3(\theta_0(u),\lambda)} d\lambda du$$

$$+ \kappa_4 \Big\{ \frac{1}{4\pi^2} \int_0^1 \Big(\int_{-\pi}^{\pi} \frac{f^2(u,\lambda)}{g^2(\theta_0(u),\lambda)} d\lambda \Big)^2 du + \frac{1}{4\pi^2} \int_0^1 \Big(\int_{-\pi}^{\pi} \frac{f(u,\lambda)}{g(\theta_0(u),\lambda)} d\lambda \Big)^2 du$$

$$- \frac{1}{2\pi^2} \int_0^1 \Big(\int_{-\pi}^{\pi} \frac{f^2(u,\lambda)}{g^2(\theta_0(u),\lambda)} d\lambda \Big) \Big(\int_{-\pi}^{\pi} \frac{f(u,\lambda)}{g(\theta_0(u),\lambda)} d\lambda \Big) du,$$

where κ_4 denotes the fourth cumulant of the innovations. If the null hypothesis (2.6) is true, we have $f(u,\lambda) = g(\theta_0(u),\lambda)$ and therefore $\sigma_1^2 = \sigma_{1,g}^2$ and $\sigma_2^2 = \sigma_{2,g}^2$.

3   Bootstrapping the test statistic

Based on Theorem 2.6 and Theorem 2.7 we have constructed tests in (2.15) and (2.17) that rely on approximations of the quantiles of the test statistics by the corresponding quantiles of the limiting standard normal distribution. However, we have experienced that the quality of these approximations is rather poor for finite samples, and it is worth noting that this is indeed the case for both situations. Thus it seems as if it is not the estimation of the variance in (2.15) that affects the accuracy of the normal approximation, but rather the shape of the finite sample distributions. For details on the precise results of the simulation study we refer to Section 4.

To improve the approximation of the nominal level we propose a bootstrap procedure to create pseudo-observations, from which in each step the test statistic is computed. By replication, we obtain empirical quantiles on which alternative tests are based. For the sake of brevity we focus on the empirical quantiles of the distribution of the statistic $\sqrt{T}\hat{G}_T/\sqrt{\hat{V}_{1,H_0,T}}$ only, but a similar bootstrap method can be derived to approximate the quantiles of the distribution of $\sqrt{2T}\hat{H}_T$.

The general procedure starts with the generation of independent identically distributed random variables $Z_1^*, ..., Z_T^* \sim N(0,1)$, from which we compute the local Fourier transforms

$$J_{N,Z}^*(u_j, \lambda_k) := \frac{1}{\sqrt{2\pi N}} \sum_{s=0}^{N-1} Z_{\lfloor u_j T \rfloor - N/2 + 1 + s, T}^* \exp(-i\lambda_k s)$$

11

for $i = 1, ..., M$ and $k = 1, ..., \lfloor \frac{N}{2} \rfloor$. In a next step we set

$$(3.1) \qquad J_{N,X}^*(u_j, \lambda_k) := \sqrt{2\pi g(\hat{\theta}_T(u_j), \lambda_k)} J_{N,Z}^*(u_j, \lambda_k),$$

where $\hat{\theta}_T(u_j)$ is the estimator of $\theta_0(u_j)$ using the original data $X_{1,T}, X_{2,T}, ..., X_{T,T}$. We then calculate a bootstrap version of the local periodogram via

$$I_N^*(u_j, \lambda_k) := |J_{N,X}^*(u_j, \lambda_k)|^2.$$

This definition makes sense, as the standard result for stationary processes, which says that the periodogram of a linear process can essentially be approximated by the product of its spectral density and the periodogram of the innovations, also applies for locally stationary ones. Finally, the test statistic for the bootstrap data is defined by $\sqrt{T}\hat{G}_T^* / \sqrt{\hat{V}_{1,T}^*}$, where

$$\hat{G}_T^* := \frac{1}{T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \left(I_N^*(u_j, \lambda_k) - g(\hat{\theta}_T^*(u_j), \lambda_k) \right)^2 - \frac{1}{2T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} I_N^*(u_j, \lambda_k)^2,$$

$$\hat{V}_{1,T}^* := \frac{1}{T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} g^4(\hat{\theta}_T^*(u_j), \lambda_k),$$

are the bootstrap analogues of the statistics $\hat{G}_T$ and $\hat{V}_{1,H_0,T}$ defined in (2.13) and (2.14), respectively, and $\hat{\theta}_T^*(u)$ is the estimator for $\theta_0(u)$ using the local periodogram $I_N^*(u, \lambda)$. Note that from (3.1) one could also obtain a pseudo-series $X_{1,T}^*, ..., X_{T,T}^*$ by the inversion formula for the discrete Fourier transformation, which is not necessary here, since we only require replications of the local periodogram to compute our test statistic.

The following result shows that under the null hypothesis the bootstrap method leads to a valid approximation of the distribution of the statistic $\hat{G}_T$. We follow Bickel and Freedman (1981) and measure the distance between distributions F and G by Mallow's distance

$$d_2(F, G) = \inf \left(E|X - Y|^2 \right)^{1/2},$$

where the infimum is taken over all pairs of random variables X and Y having the given marginal distributions F and G.

Theorem 3.1 *If the null hypothesis H_0 in (2.6), (2.12) and Assumptions 2.1 and 2.2 are satisfied, then, conditionally on $X_{1,T}, ..., X_{T,T}$, we have as $T \to \infty$*

$$d_2 \left(\frac{\sqrt{T}\hat{G}_T}{\sqrt{\hat{V}_{1,H_0,T}}}, \frac{\sqrt{T}\hat{G}_T^*}{\sqrt{\hat{V}_{1,T}^*}} \right) \to 0$$

in probability.

Finally, we obtain empirical quantiles as follows: Let $\hat{G}^*_{T,1}, \ldots, \hat{G}^*_{T,B}$ and $\hat{V}^*_{1,T,1}, \ldots, \hat{V}^*_{1,T,B}$ denote the sample from B bootstrap replications. Then we reject the null hypothesis whenever

$$(3.2) \qquad \frac{\hat{G}_T}{\sqrt{\hat{V}_{1,H_0,T}}} > \Big(\frac{\hat{G}^*}{\sqrt{\hat{V}^*}}\Big)_{T,(\lfloor(1-\alpha)B\rfloor)},$$

where $\Big(\frac{\hat{G}^*}{\sqrt{\hat{V}^*}}\Big)_{T,1}, \ldots, \Big(\frac{\hat{G}^*}{\sqrt{\hat{V}^*}}\Big)_{T,B}$ denotes the order statistic of $\frac{\hat{G}^*_{T,1}}{\sqrt{\hat{V}^*_{1,T,1}}}, \ldots, \frac{\hat{G}^*_{T,B}}{\sqrt{\hat{V}^*_{1,T,B}}}$. By Theorem 3.1 this test has asymptotic level α and is consistent, because conditionally on $X_{1,T}, \ldots, X_{T,T}$ each bootstrap statistic $\sqrt{T}\hat{G}^*_T/\sqrt{\hat{V}^*_{1,T}}$ converges to a standard normal distribution, while $\sqrt{T}\hat{G}_T/\sqrt{\hat{V}_{1,H_0,T}}$ blows up to infinity by Theorem 2.6, if the null hypothesis is not satisfied. We can use a similar bootstrap procedure to approximate the quantiles of the distribution of the statistic $\sqrt{2T}\hat{H}_T$. In fact only the final step differs, because we have to calculate $\hat{H}^*_T$ instead of $\hat{G}^*_T$ and $\hat{V}^*_{1,T}$.

Since we want to compare our approach with that of Sergides and Paparoditis (2009) in the next section and as these authors require a bootstrap procedure to obtain satisfying results as well, we also investigate the differences between the method suggested in this paper and the one proposed by these authors. Their methodology is limited to the case of testing for a tvAR(p) structure, so the null hypothesis is given by

$$(3.3) \qquad H_0 : f(u, \lambda) = \frac{\sigma^2(u)}{2\pi}\Big|\sum_{j=0}^{p} a_j(u)\exp(i\lambda j)\Big|^{-2},$$

for some appropriate functions $a_j(\cdot)$, $\sigma(\cdot)$, and the estimates for these functions are denoted by $\hat{a}(u) = (\hat{a}_1(u), \ldots, \hat{a}_p(u))^T$ and $\hat{\sigma}^2(u)$. As it was pointed out in Dahlhaus (2009), in the case of a tvAR(p) process the estimator $\hat{\theta}_T(u)$ is the same as the estimator obtained by solving the local Yule-Walker equations. Consequently, the estimate $\hat{a}(u)$ can be obtained as a solution of the equation $-\hat{S}(u)\hat{a}(u) = \hat{s}(u)$, where

$$\hat{S}(u) := \hat{c}_N(u, i-j)_{i,j=1,\ldots,p}, \quad \hat{s}(u) := (\hat{c}_N(u,1), \ldots, \hat{c}_N(u,p))^T,$$

and

$$\hat{c}_N(u,j) := \frac{1}{N}\sum_{\substack{k,l=0 \\ k-l=j}}^{N-1} X_{\lfloor uT\rfloor-N/2+k+1,T}X_{\lfloor uT\rfloor-N/2+l+1,T} = \int_{-\pi}^{\pi} I_N^X(u,\lambda)\exp(-i\lambda j)d\lambda$$

plays the role of an empirical local covariance. An estimator for the variance function of the innovations is then given by $\hat{\sigma}^2(u) = \hat{c}_N(u,0) + \hat{a}(u)^T\hat{s}(u)$.

The bootstrap procedure proposed by Sergides and Paparoditis (2009) fits a tvAR(p) model to the data $X_{1,T}, \ldots, X_{T,T}$ by solving the local Yule-Walker equations. Once we have computed those estimates at times t/T, we generate pseudo observations $X^*_{1,T}, \ldots, X^*_{T,T}$ from the fitted tvAR(p) model, i.e. from

$$X^*_{t,T} + \sum_{j=1}^{p} \hat{a}_j(t/T)X^*_{t-j,T} = \hat{\sigma}(t/T)Z^*_t.$$

Here we have set $X^*_{j,T} = X_{j,T}$ for the first p observations, and the Z^*_t are i.i.d. $N(0,1)$ distributed random variables, independent of $X_{1,T}, ..., X_{T,T}$. Using these pseudo observations we then calculate the bootstrap version of our test statistic.

4 Finite sample properties

In this section we study the finite sample properties of the asymptotic level α tests defined by (2.15) and (2.17) as well as of their bootstrap versions introduced in Section 3. We show the advantages of the latter approximations and compare our approach with the one which was recently proposed by Sergides and Paparoditis (2009). All reported results are based on 500 simulation runs and 200 bootstrap replications.

Let us start with a heuristic discussion on how to choose N [or M] in practice. In general, we recommend to choose N rather large compared to M. Apart from the assumptions in (2.12), which are required for the asymptotic theory, the main reason is that we have to minimize the bias induced by the estimation of the function $\theta(u)$. Theorem 2.4 suggests that any choice satisfying $\sqrt{T}/N \to 0$ is sufficient, but in order to obtain a negligible error for realistic sample sizes, we recommend to choose N of a considerably larger order than $\sqrt{T}$. On the other hand, as we also have to address the local structure of the time series, M may not be too small. Therefore, our experience is that in sample sizes up to $T = 1024$ the choice $M = 8$ appears to be reasonable, and these recommendations are supported by the simulation results presented in the next two sections as well.

4.1 Advantages of the bootstrap method

In this paragraph we compare the new test defined in (2.15), which is based on the quantiles of the normal distribution, with the two bootstrap tests described in Section 3. In order to demonstrate the advantages of the latter method and since we want to compare our procedure with that of Sergides and Paparoditis (2009), we exemplarily consider $AR(p)$ processes of the form

$$(4.1) \qquad X_{t,T} + 0.5\cos(4\pi t/T)X_{t-1,T} + cX_{t-2,T} = Z_t,$$

with independent, standard Gaussian distributed innovations Z_t, and using different choices for the parameter c. Other scenarios show similar results and are not depicted for the sake of brevity.

The null hypothesis is that the underlying process is a tvAR(1) process with homoscedastic innovations, i.e. $\sigma^2(u) \equiv 1$. Therefore the choice $c = 0$ in (4.1) corresponds to the null hypothesis, whereas $c = 0.3$ represents an alternative. Both the results with and without bootstrap approximations are depicted in Table 1 for various sample sizes and choices of M and N. While the Bootstrap (1)-column corresponds to the test in (3.2), the Bootstrap (2)-column corresponds to the proposal of Sergides and Paparoditis (2009) which was described in the previous paragraph.

We observe that for small sample sizes the approximation using the quantiles of the standard normal distribution is rather poor, while for the sample size $T = 1024$ the approximation becomes more accurate

			Without Bootstrap				With Bootstrap (1)				With Bootstrap (2)			
			$H_0 : c = 0$		$H_1 : c = 0.3$		$H_0 : c = 0$		$H_1 : c = 0.3$		$H_0 : c = 0$		$H_1 : c = 0.3$	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
64	8	8	.004	.022	.104	.162	.070	.104	.164	.214	.054	.100	.160	.248
128	16	8	.014	.034	.146	.216	.038	.086	.272	.346	.050	.100	.262	.350
128	8	16	.006	.030	.090	.170	.056	.078	.226	.276	.064	.116	.228	.314
256	32	8	.024	.036	.298	.398	.080	.110	.448	.588	.056	.104	.480	.576
256	16	16	.006	.036	.220	.322	.048	.092	.424	.506	.048	.106	.402	.480
256	8	32	.004	.044	.092	.200	.048	.06	.286	.332	.066	.122	.292	.396
512	64	8	.028	.050	.592	.696	.050	.100	.724	.828	.056	.138	.692	.794
512	32	16	.016	.026	.500	.628	.068	.116	.750	.818	.064	.118	.690	.778
512	16	32	.008	.032	.348	.486	.032	.072	.626	.688	.076	.112	.582	.646
512	8	64	.006	.100	.142	.314	.026	.044	.328	.372	.060	.128	.306	.396
1024	128	8	.044	.076	.852	.908	.082	.136	.906	.954	.058	.100	.882	.944
1024	64	16	.020	.024	.832	.89	.054	.122	.906	.970	.074	.136	.926	.962
1024	32	32	.004	.012	.696	.786	.054	.080	.938	.958	.052	.094	.868	.920
1024	16	64	.004	.038	.510	.648	.050	.090	.758	.802	.082	.106	.680	.754
1024	8	128	.018	.198	.192	.440	.016	.032	.368	.424	.080	.148	.332	.408

Table 1: *Rejection probabilities of the test (2.15) in the model (4.1) for different values of c using the quantiles of the standard normal distribution and of the bootstrap approximations.*

			$H_0 : \phi(x)=-0.9x \atop c=0$		$H_0 : \phi(x)=0.9x \atop c=0$		$H_1 : \phi(x)=0.9x \atop c=0.2$		$H_1 : \phi(x)=0.9x \atop c=0.3$	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%
64	8	8	.042	.090	.062	.098	.090	.146	.198	.280
128	16	8	.056	.094	.056	.096	.166	.236	.278	.348
128	8	16	.050	.116	.058	.096	.096	.156	.168	.248
256	32	8	.048	.108	.048	.094	.172	.268	.334	.456
256	16	16	.054	.102	.056	.104	.184	.248	.322	.390
256	8	32	.056	.112	.048	.098	.080	.142	.140	.198

Table 2: *Rejection probabilities of the bootstrap version of the test (2.15) in model (4.2) for different choices of the function ϕ and the parameter c using the quantiles of the bootstrap approximation.*

if we choose $M = 8$, as it was recommended in the previous paragraph. In comparison to that, the results using the quantiles of the bootstrap approximation are quite satisfying even for small sample sizes [note that $T = 64$ corresponds to a very small sample size, as we need both N and M to be large for the asymptotics to kick in]. Furthermore, the approximation of the nominal level of the bootstrap test is not very sensitive with respect to the choice of M, and we observe a slightly better performance of the bootstrap method (2). Under the alternative represented by the choice $c = 0.3$ the bootstrap tests are usually more powerful, and this can only be partially explained by the fact that the asymptotic level α test from (2.15) is very conservative. In most cases we observe a slightly larger power of the bootstrap method (1). The simulation results of the power also confirm our recommendation to choose M equal to 8, and we note that the bootstrap method is only sensitive to odd choices of N and M that clearly contradict the assumption (2.12).

For the test (2.17) and its corresponding bootstrap version we obtain similar results, which are not reported for the sake of brevity.

4.2 Size and power performance of the test

From now on, following the results of the last paragraph, we only investigate the behaviour of the tests (2.15) and (2.17) using a bootstrap approximation. Furthermore, as we are interested in a fair comparison between the [bootstrap versions of the] tests (2.15), (2.17) and the kernel-based test derived in Sergides and Paparoditis (2009) later on, we restrict ourselves to their bootstrap method. We also focus on the cases $T = 64$, $T = 128$ and $T = 256$ only, since our method seems to work well even for these small sample sizes.

As in the last section the null hypothesis is that the underlying process has a tvAR(1) structure with

T	N	M	$H_0 : \begin{smallmatrix}\phi(x)=-0.9x\\c=0\end{smallmatrix}$		$H_0 : \begin{smallmatrix}\phi(x)=0.9x\\c=0\end{smallmatrix}$		$H_1 : \begin{smallmatrix}\phi(x)=0.9x\\c=0.2\end{smallmatrix}$		$H_1 : \begin{smallmatrix}\phi(x)=0.9x\\c=0.3\end{smallmatrix}$	
			5%	10%	5%	10%	5%	10%	5%	10%
64	8	8	.028	.042	.012	.056	.024	.060	.096	.182
128	16	8	.066	.088	.022	.062	.138	.240	.420	.574
128	8	16	.024	.078	.002	.022	.008	.018	.046	.098
256	32	8	.060	.072	.038	.080	.118	.244	.670	.784
256	16	16	.042	.090	.020	.074	.075	.160	.638	.750
256	8	32	.008	.044	.002	.026	.016	.044	.034	.124

Table 3: *Rejection probabilities of the bootstrap version of the test (2.17) in model (4.2) for different choices of the function ϕ and the parameter c using the quantiles of the bootstrap approximation.*

$\sigma^2(u) \equiv 1$. We consider processes of the form

(4.2)
$$X_{t,T} + \phi(t/T)X_{t-1,T} + cX_{t-2,T} = Z_t,$$

with different functions $\phi : \mathbb{R} \to \mathbb{R}$ and different values for c. The results are depicted in Table 2 and Table 3 for the bootstrap versions of the tests (2.15) and (2.17), respectively. They confirm the results from the last paragraph [namely that $M = 8$ seems to be the best choice] and they show that in most cases the bootstrap test based on the statistic D^2 is much more robust with respect to the choice of N and M than the test based on the statistic R^2 defined in (2.10). In the latter case the differences caused by the choice of M and N can be substantial, in particular for small sample sizes. For example, if $N = 32, M = 8$ the power of the bootstrap version of the test (2.17) for $c = 0.3$ is .118 while it is .075 if $N = M = 16$. On the other hand, this test yields better results for $c = 0.3$, if $T \geq 128$ and N and M are chosen in a way that does not contradict (2.12).

We finally compare the power behaviour of the new approach to the one suggested in Sergides and Paparoditis (2009). These authors basically estimate the distance $4\pi R^2$ by a feasible version of

$$Q_T := \frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \left\{ \frac{1}{N} \sum_k K_b(\lambda - \lambda_j) \left(\frac{I_N^X(u_j, \lambda_k)}{g(\hat{\theta}_T(u_j), \lambda_k)} - 1 \right) \right\}^2 d\lambda$$

with $K_b(x) = \frac{K(x/b)}{b}$ for some kernel function K and a bandwidth b. The results of the corresponding bootstrap test are shown in Table 4 for different choices of the bandwidth b and the parameter M. It can be seen that this method depends in the same way on the choice of the parameter M as our procedure does, and the best results are also usually obtained by choosing $M = 8$. Furthermore, the method seems to be quite sensible with respect to the choice of the bandwidth b, although we considered only a rather small range of all possible values of b [while we showed results for essentially all reasonable values of N and M]. The bootstrap test based on the statistic D^2 is in most examples better or at least

			$H_1 : c = 0.2$					
			$b = 0.4$		$b = 0.5$		$b = 0.6$	
T	N	M	5%	10%	5%	10%	5%	10%
64	8	8	.045	.105	.045	.120	.035	.105
128	16	8	.134	.234	.078	.150	.058	.136
128	8	16	.056	.108	.046	.084	.034	.090
256	32	8	.238	.352	.14	.262	.136	.230
256	16	16	.132	.230	.126	.238	.066	.146
256	8	32	.034	.098	.040	.098	.036	.074
			$H_1 : c = 0.3$					
			$b = 0.4$		$b = 0.5$		$b = 0.6$	
T	N	M	5%	10%	5%	10%	5%	10%
64	8	8	.108	.190	.09	.182	.068	.152
128	16	8	.306	.424	.204	.324	.192	.308
128	8	16	.126	.242	.104	.228	.098	.152
256	32	8	.600	.710	.484	.604	.346	.496
256	16	16	.336	.514	.288	.458	.218	.344
256	8	32	.134	.252	.086	.198	.092	.208

Table 4: *Rejection probabilities of the test proposed in Sergides and Paparoditis (2009) in the model (4.2) for $\phi(x) = 0.9x$ and different choices of the parameter c.*
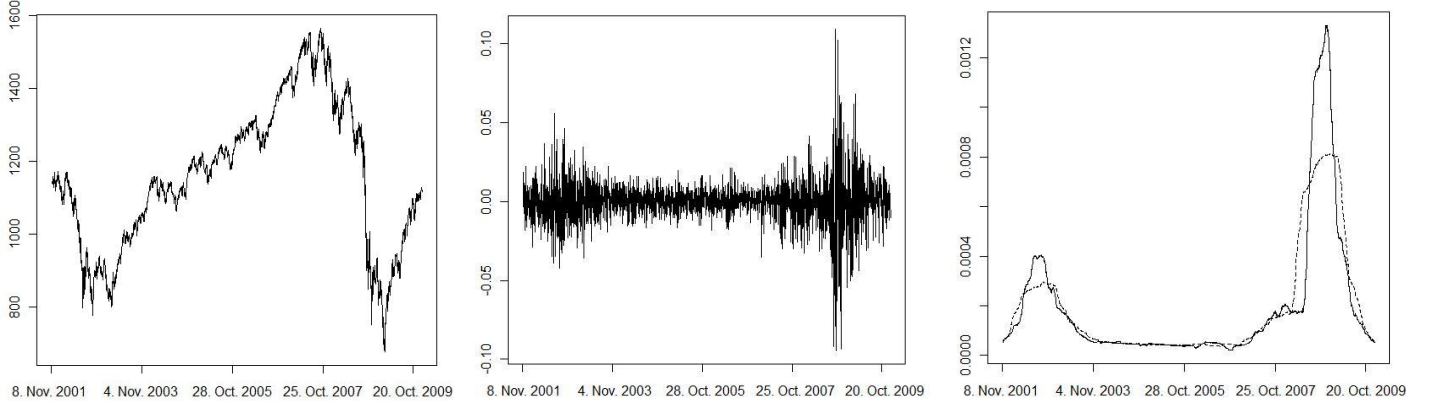
Figure 1: *The left panel displays the price of the S&P 500 index between 11/08/2001 and 12/31/2009, whereas the log returns of the S&P 500 in the same period are shown in the middle panel. Two variance estimators $\hat{\sigma}^2(\cdot)$ are depicted in the right panel. The dotted line corresponds to the choice $N = 256$, while the solid line belongs to the choice $N = 128$.*

competitive to that of Sergides and Paparoditis (2009). The approach of these authors yields better results than the test proposed in this paper for some choices of N, M and b, if $c = 0.3$ and $T \geq 128$. On the other hand, it is clearly outperformed in these cases by the bootstrap test based on the statistic R^2.

4.3 Data example

In this section we illustrate the performance of the test developed in the present paper in a data example investigating 2049 observations of the S&P 500 index which were recorded between 11/08/2001 and 12/31/2009. We consider the log returns $Y_i = \log\left(\frac{X_{i+1}}{X_i}\right)$ $(i = 1, ..., 2048)$, and the null hypothesis is that these log returns are Gaussian with a time varying variance. Precisely, we are testing if

$$(4.3) \qquad\qquad H_0: \quad Y_{i,T} = \sigma\left(\frac{i}{2048}\right)Z_i \quad \text{for } i = 1, ..., T,$$

with $Z_i \sim$ iid $N(0,1)$ and $Y_{i,T} = Y_i$ for $T = 2048$. The S&P 500 prices, the log returns and two estimators $\hat{\sigma}^2(\cdot)$ for the variance function are displayed in Figure 1. The two variance estimators were obtained by calculating $\hat{\theta}_T(i/T)$ for $i = 1, ..., 2048$ using a rolling window of either $N = 256$ (dotted line) or $N = 128$ (solid line) observations around i to compute the local periodogram $I_N^X(i/T, \lambda)$ from. Both estimates reproduce the rather large volatility in the aftermath of the dotcom crash and 9/11 and during the credit crisis 2008 very well, and the estimator using a smaller window yields a larger variance for both crashes, as might have been expected.

In order test the null hypothesis (4.3) we use the bootstrap version of the tests (2.15) and (2.17) both with $N = 256$ and $N = 128$. The p-values are depicted in Table 5, and again we note that the test

			With Bootstrap	
T	N	M	Test (2.15)	Test (2.17)
2048	256	8	.139	.223
2048	128	16	.180	.524

Table 5: *P-values of the tests* (2.15) *and* (2.17) *for the null hypothesis* (4.3) *using different choices for* M.

based on the statistic $\hat{G}_T$ is much more robust with respect to different choices of M and N. In all cases we observe p-values larger than 10%, which means that there is not enough evidence to reject the null hypothesis (4.3).

5 Appendix: Some technical details

Before we come to the proof of the main results in this paper, let us start with introducing two important auxiliary claims which have applications beyond the scope of this paper. In both cases we will only give a sketch of the proof, as a formal proof of a closely related (but slightly less general) theorem can be found in Dette et al. (2010).

Lemma 5.1 *Let* $\phi_i : [0,1] \times [-\pi, \pi] \to \mathbb{R}$, $i = 1, 2$, *be continuously differentiable functions with* $\phi_i(u, \lambda) = \phi_i(u, -\lambda)$. *If* (2.12) *is satisfied, we have*

$$\sqrt{T} \begin{pmatrix} \frac{1}{T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \phi_1(u_j, \lambda_k) I_N^X(u_j, \lambda_k)^2 - \mu_1 \\ \frac{1}{T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \phi_2(u_j, \lambda_k) I_N^X(u_j, \lambda_k) - \mu_2 \end{pmatrix} \xrightarrow{D} \mathcal{N}(0, \Sigma),$$

with mean

$$\mu_1 = \frac{1}{2\pi} \int_0^1 \int_{-\pi}^{\pi} \phi_1(u, \lambda) f^2(u, \lambda) d\lambda du$$

$$\mu_2 = \frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} \phi_2(u, \lambda) f(u, \lambda) d\lambda du$$

and variance

$$\Sigma = \begin{pmatrix} \frac{5}{\pi} \int_0^1 \int_{-\pi}^{\pi} \phi_1^2(u, \lambda) f^4(u, \lambda) d\lambda du & \frac{1}{\pi} \int_0^1 \int_{-\pi}^{\pi} \phi_1(u, \lambda) \phi_2(u, \lambda) f^3(u, \lambda) d\lambda du \\ \frac{1}{\pi} \int_0^1 \int_{-\pi}^{\pi} \phi_1(u, \lambda) \phi_2(u, \lambda) f^3(u, \lambda) d\lambda du & \frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} \phi_2^2(u, \lambda) f^2(u, \lambda) d\lambda du \end{pmatrix}.$$

20

Proof: Let us give some basic ideas only. With $t_j = u_j T$ we have

$$I_N^X(u_j, \lambda_k) = \frac{1}{2\pi N} \sum_{r,s=0}^{N-1} X_{t_j-N/2+1+r,T} X_{t_j-N/2+1+s,T} \exp(-i\lambda_k(r-s)),$$

and from (2.2) it follows that the error due to the approximation of $X_{t,T}$ by the corresponding series

(5.1) $$X_{t,T} = \sum_{l=-\infty}^{\infty} \psi_l(t/T) Z_{t-l}, \quad t = 1, \ldots, T,$$

is of small order. Using the latter representation and equations (2.3)–(2.5) one can show that a Taylor expansion of ψ in a neighbourhood of u_j yields

$$I_N^X(u_j, \lambda_k) = \frac{1}{2\pi N} \sum_{l,m=-\infty}^{\infty} \psi_l(u_j) \psi_m(u_j) \sum_{r,s=0}^{N-1} Z_{t_j-N/2+1+r-l} Z_{t_j-N/2+1+s-m} \exp(-i\lambda_k(r-s)) + O_p(\frac{N^2}{T^2}).$$

Starting from this representation all claims regarding asymptotic means and variances can be derived. For example, using (2.3)–(2.5) again one obtains

$$
\begin{aligned}
E[I_N^X(u_j, \lambda_k)] &= \frac{1}{2\pi N} \sum_{l,m=-\infty}^{\infty} \psi_l(u_j) \psi_m(u_j) \exp(-i\lambda_k(l-m)) + O(1/N) + O(N^2/T^2) \\
&= f(u_j, \lambda_k) + O(1/N) + O(N^2/T^2).
\end{aligned}
$$

Since

(5.2) $$\int_0^1 \int_{-\pi}^{\pi} \phi_2(u, \lambda) f(u, \lambda) d\lambda du = \frac{4\pi}{T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \phi_2(u_j, \lambda_k) f(u_j, \lambda_k) + O(1/N^2) + O(\frac{N^2}{T^2})$$

from the choice of u_j and by a periodic extension of ϕ_2 in its second component, (2.12) yields

$$\sqrt{T} E\left[\frac{1}{T} \sum_{j=1}^{M} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \phi_2(u_j, \lambda_k) I_N^X(u_j, \lambda_k) - \mu_2 \right] \to 0$$

easily. In a similar way one proves the claim regarding the mean of the first component. All (co)variances can be computed using the product theorem for cumulants from Brillinger (1981), and the latter result is the main key for the formal proof of the entire result. In fact, one uses the Cramér-Wold device and shows weak convergence of any linear combination to the corresponding normal distribution, which is usually done by establishing convergence of all cumulants. See Dette et al. (2010) for an explicit proof in a similar situation. □

21

Lemma 5.2 *Let $u \in [0,1]$ be fixed and let $\phi : [0,1] \times [-\pi, \pi] \to I\!R$ be continuously differentiable with $\phi(u, \lambda) = \phi(u, -\lambda)$. Then we have*

$$\sqrt{N}\left(\frac{1}{N}\sum_{k=1}^{\lfloor \frac{N}{2} \rfloor}\left(I_N(u, \lambda_k) - f(u, \lambda_k)\right)\phi(u, \lambda_k)\right) \xrightarrow{D} \mathcal{N}(0, \tau^2)$$

with

$$\tau^2 = \frac{1}{4\pi}\int_{-\pi}^{\pi}\phi^2(u, \lambda)f^2(u, \lambda)d\lambda.$$

This auxiliary result can be shown similarly to the previous one. Let us now come to the proof of the main results of this paper.

Proof of Lemma 2.3: Without loss of generality we assume $p = 1$. The mean value theorem gives

$$\frac{\partial}{\partial \theta}\mathcal{L}_T(u, \hat{\theta}_T(u)) - \frac{\partial}{\partial \theta}\mathcal{L}_T(u, \theta_0(u)) = \frac{\partial^2}{\partial \theta^2}\mathcal{L}_T(u, \tilde{\theta}_T^{(u)})(\hat{\theta}_T(u) - \theta_0(u))$$

with some appropriate $\tilde{\theta}_T^{(u)}$ satisfying $|\tilde{\theta}_T^{(u)} - \theta_0(u)| \leq |\hat{\theta}_T(u) - \theta_0(u)|$. For any fixed u the function $\mathcal{L}_T(u, \theta)$ is minimized by $\hat{\theta}_T(u)$, and using Lemma 5.2 plus the uniqueness of $\hat{\theta}_T(u)$ and $\theta_0(u)$ one can show easily that $\hat{\theta}_T(u)$ converges in probability to $\theta_0(u)$. Thus we can assume that $\hat{\theta}_T(u)$ is an interior point and obtain

(5.3) $$-\frac{\partial}{\partial \theta}\mathcal{L}_T(u, \theta_0(u)) = \frac{\partial^2}{\partial \theta^2}\mathcal{L}_T(u, \tilde{\theta}_T^{(u)})(\hat{\theta}_T(u) - \theta_0(u)).$$

A straightforward calculation yields

$$\frac{\partial}{\partial \theta}\mathcal{L}_T(u, \theta_0(u)) = \frac{1}{N}\sum_{k=1}^{\lfloor \frac{N}{2} \rfloor}\left(I_N^X(u, \lambda_k) - g(\theta_0(u), \lambda_k)\right)\frac{\partial}{\partial \theta}\left(g^{-1}(\theta_0(u), \lambda_k)\right)$$

and

$$\frac{\partial^2}{\partial \theta^2}\mathcal{L}_T(u, \theta_0(u)) = \frac{1}{N}\sum_{k=1}^{\lfloor \frac{N}{2} \rfloor}\left\{\left(I_N^X(u, \lambda_k) - g(\theta_0(u), \lambda_k)\right)\frac{\partial^2}{\partial \theta^2}\left(g^{-1}(\theta_0(u), \lambda_k)\right)\right.$$
$$\left. - \frac{\partial}{\partial \theta}g(\theta_0(u), \lambda_k)\frac{\partial}{\partial \theta}\left(g^{-1}(\theta_0(u), \lambda_k)\right)\right\}.$$

Note that

$$\frac{1}{N}\sum_{k=1}^{\lfloor \frac{N}{2} \rfloor}\left(f(u, \lambda_k) - g(\theta_0(u), \lambda_k)\right)\frac{\partial}{\partial \theta}\left(g^{-1}(\theta_0(u), \lambda_k)\right)$$
$$= \frac{1}{4\pi}\int_{-\pi}^{\pi}\left(f(u, \lambda) - g(\theta_0(u), \lambda)\right)\frac{\partial}{\partial \theta}\left(g^{-1}(\theta_0(u), \lambda)\right)d\lambda + O(1/N)$$
$$= \frac{\partial}{\partial \theta}\mathcal{L}(u, \theta_0(u)) + O(1/N) = O(1/N)$$

22

by definition of $\theta_0(u)$, thus

$$\frac{\partial}{\partial\theta}\mathcal{L}_T(u,\theta_0(u)) = \frac{1}{N}\sum_{k=1}^{\lfloor\frac{N}{2}\rfloor}\left(I_N^X(u,\lambda_k) - f(u,\lambda_k)\right)\frac{\partial}{\partial\theta}\left(g^{-1}(\theta_0(u),\lambda_k)\right) + O(\frac{1}{N}).$$

Using Lemma 5.2 we obtain

$$(5.4) \qquad\qquad\qquad \frac{\partial}{\partial\theta}\mathcal{L}_T(u,\theta_0(u)) = O_P(1/\sqrt{N}).$$

Finally,

$$(5.5) \qquad \frac{\partial^2}{\partial\theta^2}\mathcal{L}_T(u,\theta_0(u)) \xrightarrow{P} \frac{1}{4\pi}\int_{-\pi}^{\pi}\left(f(u,\lambda) - g(\theta_0(u),\lambda)\right)\frac{\partial^2}{\partial\theta^2}\left(g^{-1}(\theta_0(u),\lambda)\right)d\lambda$$
$$-\frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\partial}{\partial\theta}g(\theta_0(u),\lambda)\frac{\partial}{\partial\theta}\left(g^{-1}(\theta_0(u),\lambda)\right)d\lambda =: \Gamma(u),$$

which is positive definite, since $\theta_0(u)$ is unique. From $\tilde{\theta}_T^{(u)} \to \theta_0(u)$ in probability the claim follows. $\square$

Proof of Lemma 2.5: By definition of $\hat{F}_T^o$ we have

$$\hat{F}_T^o = \frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{\lfloor\frac{N}{2}\rfloor}I_N^X(u_j,\lambda_k)^2 - 2\frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{\lfloor\frac{N}{2}\rfloor}I_N^X(u_j,\lambda_k)g(\theta_0(u_j),\lambda_k) + \frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{\lfloor\frac{N}{2}\rfloor}g^2(\theta_0(u_j),\lambda_k).$$

The latter term satisfies

$$(5.6) \qquad \frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{\lfloor\frac{N}{2}\rfloor}g^2(\theta_0(u_j),\lambda_k) = \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}g^2(\theta_0(u),\lambda)d\lambda du + o(1/\sqrt{T})$$

from (2.12) and the same estimate as in (5.2), whereas

$$(5.7) \qquad \frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{\lfloor\frac{N}{2}\rfloor}I_N^X(u_j,\lambda_k)^2 = \frac{1}{2\pi}\int_0^1\int_{-\pi}^{\pi}f^2(u,\lambda)d\lambda du + O_p(1/\sqrt{T}),$$

$$(5.8) \qquad \frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{\lfloor\frac{N}{2}\rfloor}I_N^X(u_j,\lambda_k)g(\theta_0(u_j),\lambda_k) = \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}f(u,\lambda)g(\theta_0(u),\lambda)d\lambda du + O_p(1/\sqrt{T}).$$

using Lemma 5.1. Note that $g(\theta_0(u),\lambda)$ satisfies the necessary assumptions by differentiability of θ_0. The claim follows. $\square$

Proof of Theorem 2.6: We start with the proof of

$$\sqrt{T}(\hat{G}_T^o - D^2) \xrightarrow{D} \mathcal{N}(0,\sigma^2),$$

for which we use the same decomposition as in the previous proof. Due to (5.6) it is the convergence in (5.7) and (5.8) that drives the asymptotics, thus the claim follows from Lemma 5.1 immediately. The

entire result can then be deduced from Theorem 2.4 and (2.12) via an application of the Cauchy-Schwarz inequality. $\square$

Proof of Theorem 3.1: Because of Assumption 2.1 and 2.2 the function $g(\hat{\theta}(u), \lambda)$ satisfies the assumptions on the function $\phi(u, \lambda)$ in Lemma 5.1. Thus following the same steps as in the proof of Theorem 2.6 we see, that the limits of all cumulants of $\sqrt{T}\hat{G}_T^*$ converge to the cumulants of the limiting normal distribution given in Theorem 2.6. $\square$

References

Berger, J. O. and Delampady, M. (1987). Testing precise hypotheses. *Statistical Science*, 2:317–335.

Bickel, P. and Freedman, D. (1981). Some asymptotic theory for bootstrap. *Annals of Statistics*, 9:1196–1217.

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*. McGraw Hill, New York.

Chang, C. and Morettin, P. (1999). estimation of time-varying linear systems. *Statistical Inference for Stochastic Processes.*, 2:253–285.

Dahlhaus, R. (1996). On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Process and their Applications*, 62:139–168.

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

Dahlhaus, R. (2009). Local inference for locally stationary time series based on the empirical spectral measure. *Journal of Econometrics*, 151:101–112.

Dahlhaus, R., Neumann, M., and von Sachs, R. (1999). Nonlinear wavelet estimation of time varying autoregressive processes. *Bernoulli*, 5:873–906.

Dahlhaus, R. and Polonik, W. (2006). Nonparametric quasi maximum likelihood estimation for Gaussian locally stationary processes. *Annals of Statistics*, 34(6):2790–2824.

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

Delgado, M. A., Hidalgo, J., and Velasco, C. (2005). Distribution free goodness-of-fit tests for linear processes. *Annals of Statistics*, 33:2568–2609.

Dette, H., Preuss, P., and Vetter, M. (2010). A measure of stationarity in locally stationary processes with applications to testing. Preprint, Ruhr-University Bochum.

Dette, H. and Spreckelsen, I. (2003). A note on a specification test for time series models based on spectral density estimation. *Scandinavian Journal of Statistics*, 30:481–491.

Eichler, M. (2008). Testing nonparametric and semiparametric hypotheses in vector stationary processes. *Journal of Multivariate Analysis*, 99:968–1009.

Hjellvik, V., Yao, Q., and Tjøstheim, D. (1998). Linearity testing using local polynomial approximation. *Journal of Statistical Planning and Inference*, 68:295–321.

Neumann, M. H. and von Sachs, R. (1997). Wavelet thresholding in anisotropic function classes and applications to adaptive estimation of evolutionary spectra. *Annals of Statistics*, 25:38–76.

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

Paparoditis, E. (2000). Spectral density based goodness-of-fit tests for time series models. *Scandinavian Journal of Statistics*, 27:143–176.

Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society, Ser. B*, 62:204–237.

Sakiyama, K. and Taniguchi, M. (2003). Testing composite hypotheses for locally stationary processes. *Journal of Time Series Analysis*, 24(4):483–504.

Sakiyama, K. and Taniguchi, M. (2004). Discriminant analysis for locally stationary processes. *Journal of Multivariate Analysis*, 90:282–300.

Sergides, M. and Paparoditis, E. (2008). Bootstrapping the local periodogram of locally stationary processes. *Journal of Time Series Analysis.*, 29:264–299.

Sergides, M. and Paparoditis, E. (2009). Frequency domain tests of semiparametric hypotheses for locally stationary processe. *Scandinavian Journal of Statistics*, 36:800–821.

Van Bellegem, S. and von Sachs, R. (2008). Locally adaptive estimation of evolutionary wavelet spectra. *Annals of Statistics*, 36(4):1879–1924.