A note on a specification test for time series models based on spectral density estimation

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: Holger.Dette@ruhr-uni-bochum.de
FAX: +49 234 32 14559

Ingrid Spreckelsen
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: Ingrid.Spreckelsen@ruhr-uni-bochum.de

December 21, 2001

Abstract

In a recent paper Paparoditis (2000) proposed a new goodness-of-fit test for time series models based on spectral density estimation. The test statistic is based on the distance between a kernel estimator of the ratio of the true and the hypothesized spectral density and the expected value of the estimator under the null and provides a quantification of how well the parametric density fits the sample spectral density. In this note we give a detailed asymptotic analysis of the corresponding procedure under fixed alternatives.

Keywords and phrases: goodness-of-fit test, kernel estimator, periodogram, Whittle estimator.

1 Introduction and statement of the main result

Consider a real valued stationary stochastic process $(X_t)_{t\in\mathbb{Z}}$ with spectral density f. We are interested in the problem of testing the hypothesis that this density belongs to a certain parametric class $\mathcal{F}_\Theta$ (for example the class of densities corresponding to ARMA(p,q) processes), that is

$$H_0 : f \in \mathcal{F}_\Theta \quad \text{vs.} \quad H_1 : f \notin \mathcal{F}_\Theta. \tag{1.1}$$

Much effort has been devoted to the problem of testing parametric hypotheses for stationary processes in the time domain [see for example Ljung and Box (1978), Hong-Zhi and Bing (1991), Fan and Li (1996), Hjellvik, Yao and Tjøstheim (1998) among many others], but much less attention has been paid to the corresponding problem in the frequency domain. Recently Paparoditis (2000) proposed a test which compares the periodogram

$$I(\lambda) = \frac{1}{2\pi n}\Big|\sum_{t=1}^{n} X_t\, e^{-it\lambda}\Big|^2, \qquad \lambda \in [-\pi,\pi] \tag{1.2}$$

with an appropriate estimate of the spectral density in the class $\mathcal{F}_\Theta$. To be precise let $\mathcal{F}$ denote the class of all spectral densities which are Lipschitz continuous of order $> 1/2$ and positive on the interval $[-\pi, \pi]$. We assume that the spectral density of the data generating process $f \in \mathcal{F}$ has a corresponding Lipschitz constant $\gamma_f > 1/2$ and that the functions $f(\cdot, \theta)$ in the given class of parametric functions

$$\mathcal{F}_\Theta = \{f(\cdot,\theta) \mid f(\cdot,\theta) \in \mathcal{F} \text{ und } \theta \in \Theta\} \tag{1.3}$$

are twice continuously differentiable, where the set of parameters Θ is supposed to be compact with a nonempty interior. The Whittle estimator $\hat{\theta}_n$ [see Whittle (1953)] of the parameter θ is defined as

$$\hat{\theta}_n = \arg\min_\theta L(\theta, I) \tag{1.4}$$

where

$$L(\theta, I) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \Big\{ \log f(\lambda,\theta) + \frac{I(\lambda)}{f(\lambda,\theta)} \Big\}\, d\lambda \tag{1.5}$$

denotes the Whittle Likelihood function. Roughly speaking it follows that at the Fourier frequencies $\lambda_j = 2\pi j/n$ ($j = -N, \ldots, N = \lfloor (n-1)/2 \rfloor$) the periodogram and the parametric estimate $f(\cdot, \hat{\theta}_n)$ estimate the same object if the null hypothesis (1.1) is valid. Therefore Paparoditis (2000) proposed

$$S_n = \int_{-\pi}^{\pi} \Big\{ \frac{1}{n} \sum_{j=-N}^{N} K_h(\lambda - \lambda_j) \Big(\frac{I(\lambda_j)}{f(\lambda_j, \hat{\theta}_n)} - 1 \Big) \Big\}^2 d\lambda \tag{1.6}$$

as a test statistic for the hypothesis (1.1) where K denotes a Lipschitz continuous and symmetric kernel of order $\gamma_K > 1/2$ with support $[-\pi, \pi]$ and $K_h(\cdot) = \frac{1}{h} K(\frac{\cdot}{h})$ denotes the scaled kernel. Assume that the process $(X_t)_{t\in\mathbb{Z}}$ has a representation of the form

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \tag{1.7}$$

where

$$\sum_{j=-\infty}^{\infty} |j|^{\frac{1}{2}} |\psi_j| < \infty \tag{1.8}$$

$(\psi_0 = 1)$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a sequence of independent Gaussian variables with with zero expectation and variance $\sigma^2 > 0$. Under the assumption that the null hypothesis (1.1) is valid it was shown by Paparoditis (2000)

$$n\sqrt{h}(S_n - \mu_h) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2) \tag{1.9}$$

whenever $n \to \infty, h \sim cn^{-\delta}$ for some constant $c > 0$ and $\delta \in (0,1)$, where the asymptotic bias and variance are given by $\mu_h = n^{-1}h^{-1}\int_{-\pi}^{\pi} K^2(x)\,dx$ and

$$\tau^2 = \frac{1}{\pi} \int_{-2\pi}^{2\pi} \Big[\int_{-\pi}^{\pi} K(u)K(u+x)\,du \Big]^2 dx,$$

respectively. If we assume further that the minimizing parameter

$$\theta^* = \arg\min_{\theta\in\Theta} L(\theta, f)$$

exists and is unique, then it can be shown under additional technical assumptions that

$$S_n \xrightarrow{p} \int_{-\pi}^{\pi} D^2(\lambda)\, d\lambda \tag{1.10}$$

whenever $f \in \mathcal{F} \backslash \mathcal{F}_\Theta$, where

$$D(\lambda) = \frac{f(\lambda)}{f(\lambda, \theta^*)} - 1 \tag{1.11}$$

is a measure of deviation from the null hypothesis (1.1). Note that (1.9) and (1.10) imply the consistency of the test which rejects the hypothesis (1.1) for large values of the statistic S_n. The purpose of the present note is to further refine this statement by considering the asymptotic distributional behaviour of the statistic S_n under fixed alternatives [for local alternatives see Paparoditis (2000)]. To this end assume for the bandwidth h in the statistic (1.6)

$$h \sim n^{-\delta} \tag{1.12}$$

where $\delta = \min\{\gamma_f - 1/2, (\gamma_K - 1/2)/(\gamma_K + 1)\}$ and $\gamma_f, \gamma_K > 1/2$ denote the Lipschitz constants corresponding to the spectral density f and the kernel K, respectively.

Theorem 1.1. *If assumptions (1.7), (1.8) and (1.12) are satisfied and $f \in \mathcal{F} \backslash \mathcal{F}_\Theta$, then (as $n \to \infty$)*

$$\sqrt{n}(S_n - b_h) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \beta^2), \tag{1.13}$$

where the asymptotic bias and variance are given by

$$b_h = \int_{-\pi}^{\pi} \Big\{ \frac{1}{2\pi} \int K(u) D(\lambda - uh)\, du \Big\}^2 d\lambda,$$

$$\beta^2 = \int 16\pi \Big[(D(\lambda) + 1) D(\lambda) + \Big\{ \int \nabla \frac{1}{f(\lambda, \theta^*)} f(\lambda) D(\lambda)\, d\lambda \Big\}^{\mathsf{T}} \rho_f(\lambda) f(\lambda) \Big]^2 d\lambda$$

respectively, and

$$\rho_f(\lambda) = -\Big[\int_{-\pi}^{\pi} \{\nabla^2 \frac{1}{f(\lambda, \theta^*)} \cdot f(\lambda) + \nabla^2 \log f(\lambda, \theta^*)\}\, d\lambda \Big]^{-1} \nabla \frac{1}{f(\lambda, \theta^*)}.$$

Remark 1.2. The proof of Theorem 1.1 is tedious and outlined in the following section. The asymptotic distribution under a fixed alternative f can be used for an approximation of the power function of the test which rejects the null hypothesis in (1.2) for large values of the test statistic S_n. Alternatively, results for local asymptotics [see Paparoditis (2000)] could be used for such a power approximation. It has been demonstrated by Dette and Neumeyer (2000) in the context of testing independence that asymptotics under fixed alternatives usually provides a more accurate approximation of the power function than asymptotics under local alternatives. In the present situation the asymptotic properties under fixed alternatives are based on the classical central limit theorem (see the proof of Theorem 1.1 in Section 2) while the local asymptotic properties derived by Paparoditis (2000) require a central limit theorem for random quadratic forms. This observation provides a theoretical justification that for most alternatives Theorem 1.1 yields a more accurate approximation of the power than a corresponding statement under local alternatives.

We conclude this section with an example, namely testing randomness against the alternative models AR(1) and MA(1), respectively.

Example 1.3 Consider the null hypothesis that $(X_t)_{t\in\mathbb{Z}}$ is White Noise. Then we have $\theta = \sigma^2$, and the spectral density is given by $f(\lambda,\theta) = \frac{\theta}{2\pi}$.

(1) If the true model is AR(1), that is $X_t = \phi X_{t-1} + \varepsilon_t$ where $|\phi| < 1$, then we have $f(\lambda) = \frac{\theta}{2\pi}(1 - 2\phi\cos(\lambda) + \phi^2)^{-1}$ and the asymptotic bias and variance in Theorem 1.1 are given by

$$b = \frac{2\pi\phi^2(\phi^4 - \phi^2 - 2)}{(1-\phi^2)^3}, \quad \beta^2 = \frac{32\pi^2\phi^2(2 + 2\phi^2 + 12\phi^4 - 7\phi^6 + \phi^8)}{(1-\phi^2)^7}.$$

(2) If the true model is MA(1), that is $X_t = a\varepsilon_{t-1} + \varepsilon_t$, then we have $f(\lambda) = \frac{\theta}{2\pi}(1 + 2a\cos(\lambda) + a^2)$ and the asymptotic bias and variance are given by

$$b = 2\pi a^2(2 + a^2), \quad \beta^2 = \frac{32\pi^2 a^2(2 + 6a^2 + 22a^4 + 29a^6 + 24a^8 + a^{10})}{(1+2a^2)^2}.$$

2 Proofs

Proof of Theorem 1.1. The proof of Theorem 1.1 is tedious and we only sketch the main steps for the sake of brevity. For more details we refer to Spreckelsen (2002).

We begin with a derivation of an appropriate representation of the Whittle estimator. To this end let

$$I_\varepsilon(\lambda) = \frac{1}{2\pi n}\Big|\sum_{t=1}^{n} \varepsilon_t e^{-t\lambda}\Big|^2$$

denote the periodogram of the innovation process $(\varepsilon_t)_{t\in\mathbb{Z}}$ in (1.7) and define

$$V(\lambda_i) = \frac{2\pi}{\sigma^2} I_\varepsilon(\lambda_i),$$

where $\lambda_i = 2\pi i/n; i = -N,\dots,N; N = \lfloor (n-1)/2 \rfloor$. Note that the random variables $V(\lambda_i) = V(-\lambda_i)$ are exponentially distributed and independent, whenever $i \neq 0$ (because of the Gaussian innovations). From Theorem 3 in Taniguchi (1975) and an appropriate approximation of the integral by a sum we obtain the following representation

$$\hat{\theta}_n - \theta^* = \frac{1}{n}\sum_{i=-N}^{N} 2\pi\rho_f(\lambda_i) f(\lambda_i)(V(\lambda_i) - 1) + o_p(n^{-\frac{1}{2}}). \tag{2.1}$$

Recalling the definition of the test statistic in (1.6) we obtain the decomposition

$$S_n = (S_{1n} + 2S_{2n} + S_{3n}) + (2S_{4n} + 2S_{5n} + S_{6n}) + (2S_{7n} + 2S_{8n} + 2S_{9n} + S_{10n}) \tag{2.2}$$

where

$$S_{1n} = \frac{1}{n^2}\sum_{i,j=-N}^{N}\int_{-\pi}^{\pi} K_h(\lambda-\lambda_i)K_h(\lambda-\lambda_j)\Big(\frac{I(\lambda_i)}{f(\lambda_i)} - 1\Big)\Big(\frac{I(\lambda_j)}{f(\lambda_j)} - 1\Big)\, d\lambda$$

$$S_{2n} = \frac{1}{n^2}\sum_{i,j=-N}^{N}\int_{-\pi}^{\pi} K_h(\lambda-\lambda_i)K_h(\lambda-\lambda_j)\Big(\frac{I(\lambda_i)}{f(\lambda_i)}-1\Big)\frac{I(\lambda_j)}{f(\lambda_j)}(D(\lambda_j)+1)d(\lambda_j)\,d\lambda$$

$$S_{3n} = \frac{1}{n^2}\sum_{i,j=-N}^{N}\int_{-\pi}^{\pi} K_h(\lambda-\lambda_i)K_h(\lambda-\lambda_j)\frac{I(\lambda_i)}{f(\lambda_i)}(D(\lambda_i)+1)d(\lambda_i)\frac{I(\lambda_j)}{f(\lambda_j)}(D(\lambda_j)+1)d(\lambda_j)\,d\lambda$$

$$S_{4n} = \frac{1}{n^2}\sum_{i,j=-N}^{N}\int_{-\pi}^{\pi} K_h(\lambda-\lambda_i)K_h(\lambda-\lambda_j)\Big(\frac{I(\lambda_i)}{f(\lambda_i)}-1\Big)\Big(\frac{I(\lambda_j)}{f(\lambda_j)}-1\Big)D(\lambda_j)\,d\lambda$$

$$S_{5n} = \frac{1}{n^2}\sum_{i,j=-N}^{N}\int_{-\pi}^{\pi} K_h(\lambda-\lambda_i)K_h(\lambda-\lambda_j)\frac{I(\lambda_i)}{f(\lambda_i)}(D(\lambda_i)+1)d(\lambda_i)\Big(\frac{I(\lambda_j)}{f(\lambda_j)}-1\Big)D(\lambda_j)\,d\lambda$$

$$S_{6n} = \frac{1}{n^2}\sum_{i,j=-N}^{N}\int_{-\pi}^{\pi} K_h(\lambda-\lambda_i)K_h(\lambda-\lambda_j)\Big(\frac{I(\lambda_i)}{f(\lambda_i)}-1\Big)D(\lambda_i)\Big(\frac{I(\lambda_j)}{f(\lambda_j)}-1\Big)D(\lambda_j)\,d\lambda$$

$$S_{7n} = \frac{1}{n^2}\sum_{i,j=-N}^{N}\int_{-\pi}^{\pi} K_h(\lambda-\lambda_i)K_h(\lambda-\lambda_j)\Big(\frac{I(\lambda_i)}{f(\lambda_i)}-1\Big)D(\lambda_j)\,d\lambda$$

$$S_{8n} = \frac{1}{n^2}\sum_{i,j=-N}^{N}\int_{-\pi}^{\pi} K_h(\lambda-\lambda_i)K_h(\lambda-\lambda_j)\frac{I(\lambda_i)}{f(\lambda_i)}(D(\lambda_i)+1)d(\lambda_i)D(\lambda_j)\,d\lambda$$

$$S_{9n} = \frac{1}{n^2}\sum_{i,j=-N}^{N}\int_{-\pi}^{\pi} K_h(\lambda-\lambda_i)K_h(\lambda-\lambda_j)\Big(\frac{I(\lambda_i)}{f(\lambda_i)}-1\Big)D(\lambda_i)D(\lambda_j)\,d\lambda$$

$$S_{10n} = \frac{1}{n^2}\sum_{i,j=-N}^{N}\int_{-\pi}^{\pi} K_h(\lambda-\lambda_i)K_h(\lambda-\lambda_j)D(\lambda_i)D(\lambda_j)\,d\lambda$$

and where we have used the notation (1.11) and

$$d(\lambda) = \frac{f(\lambda,\theta^*)}{f(\lambda,\hat\theta_n)} - 1, \tag{2.3}$$

The terms S_{1n}, S_{2n}, S_{3n} were treated by Paparoditis (2000) who showed that

$$S_{1n} + 2S_{2n} + S_{3n} = \frac{1}{nh}\int K^2(u)\,du + o_p(n^{-\frac{1}{2}}) = o_p(n^{-1/2}). \tag{2.4}$$

The terms S_{4n}, S_{5n}, S_{6n} differ from these quantities only by bounded factors and can be treated similarly, which gives

$$\begin{aligned} 2S_{4n} + 2S_{5n} + S_{6n} &= \frac{2}{nh}\int K^2(u)\,du\int D(\lambda)\,d\lambda \\ &+ \frac{1}{nh}\int K^2(u)\,du\int D^2(\lambda)d\lambda + o_p(n^{-\frac{1}{2}}) = o_p(n^{-\frac{1}{2}}). \end{aligned} \tag{2.5}$$

Finally, a straightforward calculation shows that

$$S_{10n} = \int_{-\pi}^{\pi}\Big\{\frac{1}{2\pi}\int K(u)D(\lambda-uh)\,du\Big\}^2 d\lambda + o(n^{-\frac{1}{2}}) \tag{2.6}$$

and therefore the assertion can be proved by showing

$$2\sqrt{n}(S_{7n}+S_{8n}+S_{9n}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\beta^2). \tag{2.7}$$

To this end we use two lemmata, which will be proved at the end of this section.

Lemma 2.1. *Under the assumptions of Theorem 1.1 we have*

$$S_{7n}+S_{9n}=\frac{1}{n}\sum_{i=-N}^{N} 2\pi(D(\lambda_i)+1)D(\lambda_i)(V(\lambda_i)-1)+o_p(n^{-\frac{1}{2}}). \tag{2.8}$$

Lemma 2.2. *Under the assumptions of Theorem 1.1 we have*

$$S_{8n}=\frac{1}{n}\sum_{i=-N}^{N} 2\pi\kappa^{\mathsf{T}}\rho_f(\lambda_i)f(\lambda_i)(V(\lambda_i)-1)+o_p(n^{-\frac{1}{2}}) \tag{2.9}$$

where

$$\kappa=\int \nabla\frac{1}{f(\lambda,\theta^*)}f(\lambda)D(\lambda)\,d\lambda.$$

With the aid of Lemma 2.1 and 2.2 we obtain

$$S_{7n}+S_{8n}+S_{9n}=\frac{1}{n}\sum_{i=-N}^{N} c_i(V(\lambda_i)-1)+o_p(n^{-\frac{1}{2}}) \tag{2.10}$$

where

$$c_i=2\pi[(D(\lambda_i)+1)D(\lambda_i)+\kappa^{\mathsf{T}}\rho_f(\lambda_i)f(\lambda_i)]$$

and the random variables $V(\lambda_i)=V(-\lambda_i)$ are independent and exponentially distributed with mean 1 for $i\neq 0$. The assertion (2.7) now follows from the central limit theorem and Theorem 1.1 is a consequence of the estimates (2.4) (2.5), (2.6) and the representation (2.2). □

For the proof of Lemma 2.1 and 2.2 we need a further auxiliary result which refines Lemma 1 in Paparoditis (2000).

Lemma 2.3. *Under the assumptions of Theorem 1.1 we have for the process* $\{R(\lambda)\}_{\lambda\in[-\pi,\pi]}$ *defined by*

$$R(\lambda)=I(\lambda)-V(\lambda)f(\lambda)=I(\lambda)-\frac{2\pi}{\sigma^2}I_\varepsilon(\lambda)$$

(i) $ER(\lambda)=o(n^{-\frac{1}{2}})$ *uniformly with respect to* $\lambda\in[-\pi,\pi]$.

(ii) $\operatorname{Cov}(R(\lambda_j),R(\lambda_m))=\begin{cases} O(n^{-1}) & \text{if } \lambda_j=\pm\lambda_m \\ o(n^{-1}) & \text{if } \lambda_j\neq\pm\lambda_m. \end{cases}$

Proof of Lemma 2.3. Let $\psi(z)=\sum_{j=-\infty}^{\infty}\psi_j z^j$, then it follows from standard arguments [see e.g. Brockwell and Davis (1991)] that

$$\begin{aligned} R(\lambda) &= (2\pi n)^{-1}\psi(e^{-i\lambda})\sum_{s=1}^{n}\varepsilon_s e^{-i\lambda s}\sum_{j=-\infty}^{\infty}\psi_j e^{i\lambda j}\Big[\sum_{t=1-j}^{n-j}\varepsilon_t e^{i\lambda t}-\sum_{t=1}^{n}\varepsilon_t e^{i\lambda t}\Big] \\ &+ (2\pi n)^{-1}\psi(e^{i\lambda})\sum_{s=1}^{n}\varepsilon_s e^{i\lambda s}\sum_{j=-\infty}^{\infty}\psi_j e^{-i\lambda j}\Big[\sum_{t=1-j}^{n-j}\varepsilon_t e^{-i\lambda t}-\sum_{t=1}^{n}\varepsilon_t e^{-i\lambda t}\Big]+|Y_n(\lambda)|^2 \end{aligned} \tag{2.11}$$

where

$$(2.12) \qquad Y_n(\lambda) = (2\pi n)^{-\frac{1}{2}} \sum_{j=-\infty}^{\infty} \psi_j e^{-i\lambda j} \Big\{ \sum_{t=1-j}^{n-j} \varepsilon_t e^{-i\lambda t} - \sum_{t=1}^{n} \varepsilon_t e^{-i\lambda t} \Big\}.$$

Now an analogous argument as given in Brockwell and Davis (1991) p. 347 shows that $E|Y_n(\lambda)|^4 = O(n^{-2})$ uniformly with respect to $\lambda \in [-\pi, \pi]$. The expectations of the first two terms in (2.11) are estimated similarly and we restrict ourselves to the first term. Observing (1.8) and Lebesgue's theorem it follows that

$$\begin{aligned}
&\Big| E\Big(n^{-1} \psi(e^{-i\lambda}) \sum_{s=1}^{n} \varepsilon_s e^{-i\lambda s} \sum_{j=-\infty}^{\infty} \psi_j e^{i\lambda j} \Big[\sum_{t=1-j}^{n-j} \varepsilon_t e^{i\lambda t} - \sum_{t=1}^{n} \varepsilon_t e^{i\lambda t} \Big] \Big) \Big| \\
&= \Big| - n^{-1} \psi(e^{-i\lambda}) \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j e^{i\lambda j} \min\{|j|, n\} \Big| \\
&\leq n^{-\frac{1}{2}} |\psi(e^{-i\lambda})| \sigma^2 \sum_{j=-\infty}^{\infty} |\psi_j| \min\Big\{ \frac{|j|}{\sqrt{n}}, \sqrt{n} \Big\} \;=\; o(n^{-1/2}).
\end{aligned}$$

which proves the assertion (i) of Lemma 2.3. The second part is proved similarly and therefore omitted [see Spreckelsen (2002) for more details]. □

Proof of Lemma 2.1. Observing the representation

$$(2.13) \qquad I(\lambda) = |\psi(e^{-i\lambda})|^2 I_\varepsilon(\lambda) + R(\lambda) = V(\lambda) f(\lambda) + R(\lambda)$$

we obtain

$$(2.14) \quad S_{7n} + S_{9n} = \frac{1}{n^2} \sum_{i,j=-N}^{N} \int K_h(\lambda - \lambda_i) K_h(\lambda - \lambda_j) (V(\lambda_i) - 1)(D(\lambda_i) + 1) D(\lambda_j)\, d\lambda \;+\; \mathcal{R}_n$$

where the remainder is defined by

$$(2.15) \qquad \mathcal{R}_n = \frac{1}{n^2} \sum_{i,j=-N}^{N} \int K_h(\lambda - \lambda_i) K_h(\lambda - \lambda_j) \frac{R(\lambda_i)}{f(\lambda_i)} (D(\lambda_i) + 1) D(\lambda_j)\, d\lambda.$$

A straightforward calculation with an application of Lemma 2.3 shows

$$\begin{aligned}
E\mathcal{R}_n^2 \;&=\; \frac{2}{n^4} \sum_{i,j,l=-N}^{N} \int\!\!\int K_h(\lambda - \lambda_i) K_h(\lambda - \lambda_j) K_h(\lambda' - \lambda_i) K_h(\lambda' - \lambda_l)\, d\lambda\, d\lambda' \\
&\times\; (D(\lambda_i) + 1)^2 D(\lambda_j) D(\lambda_l) E\Big(\frac{R(\lambda_i)}{f(\lambda_i)} \Big)^2 \\
&+\; \frac{1}{n^4} \sum_{\substack{i,j,k,l=-N \\ i \neq k, i \neq -k}}^{N} \int\!\!\int K_h(\lambda - \lambda_i) K_h(\lambda - \lambda_j) K_h(\lambda' - \lambda_k) K_h(\lambda' - \lambda_l)\, d\lambda\, d\lambda' \\
&\times\; (D(\lambda_i) + 1)(D(\lambda_k) + 1) D(\lambda_j) D(\lambda_l) E\Big(\frac{R(\lambda_i)}{f(\lambda_i)} \frac{R(\lambda_k)}{f(\lambda_k)} \Big) \;=\; o(n^{-1})
\end{aligned}$$

which implies $\mathcal{R}_n = o_p(n^{-1/2})$ by Markov's inequality. Consequently it follows from (2.14) that it is sufficient to prove the representation of Lemma 2.1 for the term

$$\begin{aligned} U_n &= \frac{1}{n^2}\sum_{i,j=-N}^{N}\int K_h(\lambda-\lambda_i)K_h(\lambda-\lambda_j)(V(\lambda_i)-1)(D(\lambda_i)+1)D(\lambda_j)\,d\lambda \\ &= S_{7n}+S_{9n}+o_p(n^{-\frac{1}{2}}). \end{aligned}$$

To this end note that

$$E\Big[U_n - \frac{1}{n}\sum_{i=-N}^{N}2\pi(D(\lambda_i)+1)D(\lambda_i)(V(\lambda_i)-1)\Big]^2 = \frac{1}{n^2}\sum_{i=-N}^{N}b_i^2(D(\lambda_i)+1)^2 \;=\; o(n^{-1})$$

where the last estimate follows from the Lipschitz continuity of K and f, which implies for the coefficients

$$\begin{aligned} b_i &= \int K_h(\lambda-\lambda_i)\Big[\frac{1}{n}\sum_{j=-N}^{N}K_h(\lambda-\lambda_j)D(\lambda_j)-D(\lambda)\Big]\,d\lambda \\ &\quad + \int K_h(\lambda-\lambda_i)D(\lambda)\,d\lambda - 2\pi D(\lambda_i) \;=\; o(1) \end{aligned} \tag{2.16}$$

the estimate $b_i^2 = o(1)$ uniformly with respect to $i \in \{-N,\dots,N\}$. This completes the proof of Lemma 2.1. □

Proof of Lemma 2.2. By a Taylor expansion there exists $\tilde{\theta}_n \in \Theta$ such that $\|\tilde{\theta}_n - \theta^*\| \le \|\hat{\theta}_n - \theta^*\|$ and we have

$$\begin{aligned} S_{8n} &= \frac{1}{n^2}\sum_{i,j=-N}^{N}\int K_h(\lambda-\lambda_i)K_h(\lambda-\lambda_j)\frac{I(\lambda_i)}{f(\lambda_i)}(D(\lambda_i)+1)d(\lambda_i)D(\lambda_j)\,d\lambda \\ &= (\hat{\theta}_n-\theta^*)^{\mathsf{T}}\frac{1}{n^2}\sum_{i,j=-N}^{N}f(\lambda_i,\theta^*)\nabla\frac{1}{f(\lambda_i,\theta^*)}(D(\lambda_i)+1)\int K_h(\lambda-\lambda_i)K_h(\lambda-\lambda_j)D(\lambda_j)\,d\lambda \\ &+ (\hat{\theta}_n-\theta^*)^{\mathsf{T}}\frac{1}{n^2}\sum_{i,j=-N}^{N}f(\lambda_i,\theta^*)\nabla\frac{1}{f(\lambda_i,\theta^*)}\Big(\frac{I(\lambda_i)}{f(\lambda_i)}-1\Big)(D(\lambda_i)+1) \\ &\times \int K_h(\lambda-\lambda_i)K_h(\lambda-\lambda_j)D(\lambda_j)\,d\lambda \\ &+ (\hat{\theta}_n-\theta^*)^{\mathsf{T}}\frac{1}{2n^2}\sum_{i,j=-N}^{N}\int K_h(\lambda-\lambda_i)K_h(\lambda-\lambda_j)\frac{I(\lambda_i)}{f(\lambda_i)}(D(\lambda_i)+1)D(\lambda_j) \\ &\times f(\lambda_i,\theta^*)\nabla^2\frac{1}{f(\lambda_i,\tilde{\theta}_n)}D(\lambda_j)\,d\lambda\;(\hat{\theta}_n-\theta^*) \\ &= \frac{1}{n^2}\sum_{i,j=-N}^{N}\int K_h(\lambda-\lambda_i)K_h(\lambda-\lambda_j)\frac{I(\lambda_i)}{f(\lambda_i)}(D(\lambda_i)+1)d(\lambda_i)D(\lambda_j)\,d\lambda + o_p(n^{-1/2}), \end{aligned}$$

where the last estimate follows by a straightforward calculation of the second moment of the terms $\frac{1}{n^2}\sum_{ij}\dots$ in the second and third expression in the above decomposition observing that the periodogram satisfies

$$\mathrm{Cov}(I_X(\lambda_j), I_X(\lambda_k)) = \begin{cases} 2f^2(\lambda_j) + O(n^{-\frac{1}{2}}) & \text{if } \lambda_j = \lambda_k = 0, \\ f^2(\lambda_j) + O(n^{-\frac{1}{2}}) & \text{if } \lambda_j = \lambda_k \neq 0, \\ O(n^{-1}) & \text{if } \lambda_j \neq \pm\lambda_k, \end{cases}$$

uniformly with respect to j and $k \in \{-N, \ldots, N\}$. The assertion of Lemma 2.2 is now obtained by an approximation of the sum in the first term by an appropriate integral, that is

$$\frac{1}{n^2} \sum_{i,j=-N}^{N} f(\lambda_i, \theta^*) \nabla \frac{1}{f(\lambda_i, \theta^*)} (D(\lambda_i) + 1) \int K_h(\lambda - \lambda_i) K_h(\lambda - \lambda_j) D(\lambda_j)\, d\lambda$$
$$= \int f(\lambda, \theta^*) \nabla \frac{1}{f(\lambda, \theta^*)} (D(\lambda) + 1) D(\lambda)\, d\lambda + O(n^{-\gamma_f}) + o(1) \ = \ \kappa + o(1).$$

□

Acknowledgements. The authors would like to thank R. Dahlhaus for some help with the references and I. Gottschlich who typed most parts of this paper with considerable technical expertise. The work of H. Dette was supported by the Deutsche Forschungsgemeinschaft (SFB 475; Komplexitätsreduktion in multivariaten Datenstrukturen).

References

Brockwell, P.J., Davis, R.A. (1991). *Times Series: Theory and Methods.* Springer, N.Y.

Dette, H., Neumeyer, N. (2000). A note on a specification test of independence. *Metrika* 51, 133-142.

Fan, Y., Li, Q. (1999). Central limit theorem for degenerate U-Statistics of absolutely regular processes with applications to model specification testing. *Nonpar. Statistics* 10, 245-271.

Hjellvik, V., Yao, Q., Tjøstheim, D. (1998). Linearity testing using local polynomial approximation. *J. Stat. Plann. Inference* 68, No.2, 295-321.

Hong-zhi, A., Bing, C. (1991). A Kolmogorov Smirnov type statistic with application to test of nonlinearity in time series. *Int. Stat. Review* 59, 3, 287-307.

Ljung, G.M., Box, G.E.P. (1978). On measures of lack-of-fit in time series models. *Biometrika* 65, 297-303.

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

Spreckelsen, I. (2002). Anpassungstests für Zeitreihenmodelle im Zeit- und Frequenzbereich. *PhD Thesis*, Ruhr-Universität Bochum (in German).

Taniguchi, M. (1979). On estimation of parameters of Gaussian stationary processes. *J. Appl. Prob.* 16, 575-591.

Whittle, P. (1953). Estimation and information in stationary time series. *Ark. Mat.* 2, 423-434.