Spectral Domain Bootstrap Tests for Stationary Time Series

by G.Hainz and R.Dahlhaus

Universität Heidelberg[1]

Abstract

For stationary linear processes Kolmogorov-Smirnov type goodness-of-fit tests for compound hypotheses based on frequency domain bootstrap methods are proposed. Similar bootstrap tests for comparing the spectral distributions of two time series are suggested. The small sample performance of the tests is investigated by simulations, and a real data example is given for illustration.

1 Introduction

In time series analysis, and generally in statistics, often parametric model classes are used for purposes like prediction or (parametric) spectral density estimation, for which e.g. an autoregressive model can be applied. In this case a model class has to be selected, and then a model within the class must be estimated by some suitable estimator. There are several possibilities for model class selection, especially selection procedures based on criteria like AIC (Akaike, 1972). With these methods one tries to find the model class best suited to the particular purpose from a given set of competing model classes. However, even if we knew which of the model classes under consideration is the best, we still do not know whether it has a good fit, or whether it is just the best under several unsuitable classes.

In this paper we will deal with some goodness-of-fit test statistics for stationary linear time series which are functions of the periodogram; in contrast to most goodness-of-fit tests we test a compound hypothesis. Since the asymptotic distributions of the statistics cannot be evaluated analytically we estimate them by a bootstrap method that works on the periodogram values, which is both an obvious and a universal approach, as the method is quite independent of the model class under consideration (in contrast to similar work by Chen and Romano, 1997, who use a time domain bootstrap).

Suppose we have a class of parametric models with parameter set Θ and some observations $X_1, \ldots, X_T$ from a realisation of a stationary time series $\{X_t\}$. A goodness-of-fit

[1]This work has been supported by the Deutsche Forschungsgemeinschaft (Da 187/8-1)

AMS 1991 subject classifications. Primary 62 M 10; secondary 62 G 09.

Keywords and phrases. Goodness-of-fit test; bootstrap; wild bootstrap; comparison tests.

test can be used to check the adequacy of the model class, i.e. to test the hypothesis that the time series is generated by a model with some parameter $\theta_0 \in \Theta$. In contrast to many well-known goodness-of-fit tests (see Priestley, 1981, Ch. 6.2.6, Anderson, 1993) this is a compound hypothesis, where θ_0, the parameter which specifies the best model of the model class, is unknown and must be estimated.

We will restrict ourselves to second order properties, so it is enough to test the hypothesis

$$H_0 : f \in \Theta \text{ against } H_1 : f \notin \Theta$$

(in the sense: $H_0 : \exists \theta \in \Theta : f = f_\theta$ against $H_1 : \forall \theta \in \Theta : f \neq f_\theta$), where f and f_θ denote the spectral densities of the process and of the model with parameter θ, respectively. Dahlhaus (1988), ex. 3.4, considers the asymptotic distribution of the test statistic

$$W_T = \sup_{\lambda \in [0,\pi]} |\sqrt{T}(F_T(\lambda) - F_{\hat{\theta}}(\lambda))|,$$

where $F_T(\lambda) = \int_0^\lambda I_T(\alpha) d\alpha$ and $F_{\hat{\theta}}(\lambda) = \int_0^\lambda f_{\hat{\theta}}(\alpha) d\alpha$. Here $\hat{\theta}$ denotes the Whittle estimator of the optimal parameter $\theta_0 \in \Theta$, I_T is the periodogram of the data, and $f_{\hat{\theta}}$ is the spectral density of the estimated model. The limit distribution of W_T is difficult to calculate explicitly. Furthermore, it depends on f and f_{θ_0} which are unknown.

Another possible test statistic is Bartlett's U_p - statistic (Bartlett, 1954) generalized to the situation of a compound hypothesis

$$V_T = \sup_{\lambda \in [0,\pi]} |\sqrt{T}(\frac{F_T(\lambda)}{F_T(\pi)} - \frac{F_{\hat{\theta}}(\lambda)}{F_{\hat{\theta}}(\pi)})|.$$

In Section 2 we show that the asymptotic distribution of the test statistic V_T can be approximated by the distribution of the supremum of a certain stochastic approximation of $|\sqrt{T}(\frac{F_T(\lambda)}{F_T(\pi)} - \frac{F_{\hat{\theta}}(\lambda)}{F_{\hat{\theta}}(\pi)})|$ over all $\lambda \in (0, \pi]$. In Section 3 we find that a bootstrap of the periodogram values can imitate this distribution, thus yielding the critical values for the test. An appropriate bootstrap method is suggested which creates a test of adequate power for misspecified model classes, i.e. for model classes which do not contain the true process.

The goodness-of-fit test based on the test statistic V_T has for example no power in the case that the parameter set Θ does not contain the true innovation variance (it cancels out in the statistic V_T) and the shape of the spectral density is fitted correctly. Therefore it is shown in Section 4 that it is possible to get a similar goodness-of-fit test using the test statistic W_T. For this purpose we need a special periodogram bootstrap which also emulates the dependence structure of the periodogram (cp. Janas and Dahlhaus, 1994).

Further we consider the problem of comparing two independent time series. For the comparison one may choose a model class, estimate the necessary parameters for each of them separately, and compare the results by testing equality of the parameters. However, a more general approach is the calculation of a statistic which directly compares the spectral densities of both series without making assumptions on a parametric model for the processes (cf. Diggle and Fisher, 1991). In Section 5 bootstrap tests are introduced to compare the spectral densities of the time series, using essentially the same bootstrap methods as in Sections 3 and 4.

Some of the tests are illustrated by simulations and a real data example (Beveridge Wheat Price Index) in Section 6; some possible improvements of the methods for small sample sizes are summarized there, too. In this paper we restrict ourselves to test statistics of Kolmogorov-Smirnov type, but the method allows the application of other test statistics (e.g. Cramér-von Mises statistics), too.

Bootstrap methods for goodness-of-fit tests for time series are also investigated e.g. by Paparoditis (1995), but with a different test statistic, and by Chen and Romano (1997) with a time domain bootstrap. For a goodness-of-fit test for autoregressive models see Anderson (1997).

2 Distribution of the test statistic

The Kullback-Leibler distance for Gaussian processes can be written up to a constant as

$$\Delta(\theta) = \frac{1}{2\pi}\int_0^{\pi}(\log f_\theta(\lambda) + \frac{f(\lambda)}{f_\theta(\lambda)})d\lambda$$

(cf. Parzen, 1983); it is used as a measure for the difference between the true process and the model with parameter $\theta \in \Theta$, even for non-Gaussian processes, where $f(\lambda)$ and $f_\theta(\lambda)$ are the corresponding spectral densities. The model with parameter $\theta_0 = \operatorname{argmin}_{\theta\in\Theta}\Delta(\theta)$ is the model in Θ which approximates the true process best (in the Kullback-Leibler sense). Furthermore $f_{\theta_0} = f$ if $f \in \Theta$.

The unknown spectral density $f(\lambda)$ is approximately the expectation of the periodogram of the data $X_1, \ldots, X_T$ at frequency λ,

$$I_T(\lambda) \quad := \quad (2\pi H_{2,T})^{-1} d_T(\lambda) d_T(-\lambda),$$

where

$$d_T(\lambda) \quad := \quad \sum_{t=1}^{T} h_t X_t \exp(-i\lambda t) \tag{2.1}$$

is the Fourier transform of the tapered data, and h_t is a data taper with $H_{i,T} := \sum_{t=1}^{T} h_t^i$ (see assumption 5). The empirical Kullback-Leibler distance

$$\hat{\Delta}(\theta) \quad = \quad \frac{1}{2\pi} \int_0^{\pi} (\log f_\theta(\lambda) + \frac{I_T(\lambda)}{f_\theta(\lambda)}) d\lambda$$

is a consistent estimator of the Kullback-Leibler distance $\Delta(\theta)$. By minimizing it we get the Whittle estimate $\hat{\theta} = \operatorname{argmin}_{\theta \in \Theta} \hat{\Delta}(\theta)$ as an estimate of θ_0, which for autoregressive models is identical to the well-known Yule-Walker estimate. As far as the asymptotic convergence is concerned, all integrals can be replaced by sums, e.g. $\frac{\pi}{n} \sum_{j=1}^{n} \phi(\frac{2\pi j}{T}) I_T(\frac{2\pi j}{T})$, $n = [T/2]$, instead of $\int_0^{\pi} \phi(\lambda) I_T(\lambda) d\lambda$ (see Brillinger, 1981, Th. 5.10.2, Dahlhaus, 1985, Section 3).

Assumptions:

1. The spectral densities of the models have the form

$$f_\theta(\lambda) = \frac{\sigma^2}{2\pi} h_\tau(\lambda), \qquad \theta \in \Theta,$$

   with parameters $\theta = (\sigma^2, \tau^\top)^\top$, $\tau^\top = (\tau_1, \ldots, \tau_p)$, and fulfill the Kolmogorov equation

$$\frac{1}{\pi} \int_{\Pi} \log f_\theta(\lambda) d\lambda = \log \frac{\sigma^2}{2\pi}.$$

   Here and in the following we set $\Pi := (0, \pi]$. Θ is assumed to be compact.

2. There is a unique $\theta_0 = (\sigma_0^2, \tau_0^\top)^\top$ minimizing the Kullback-Leibler distance $\Delta(\cdot)$ in Θ, and it lies in the interior of Θ.

3. The model spectral densities $f_\theta(\lambda)$ are twice continuously differentiable as functions of $\theta \in \Theta$ with derivatives bounded uniformly for all θ and λ, and with uniformly bounded total variations as functions of λ. The matrix $\Gamma_0 = \nabla_\theta^2 \Delta(\theta_0)$ is positive definite. Further there is a c_0 such that $0 < c_0 < f_\theta(\lambda)$ for all $\lambda \in \Pi$ and all $\theta \in \Theta$.

4. $\{X_t\}$ is a linear real valued stationary process $X_t = \sum_{v=0}^{\infty} a_v \epsilon_{t-v}$, where the ϵ_t are independent identically distributed random variables with mean 0, variance σ^2 and existing cumulants of all orders κ_k, where $|\kappa_k| \leq c^k$ for some $c > 0$. Furthermore, we assume $a_v = O(|v|^{-1.5-\gamma})$ for some $\gamma > 0$, and $f(\lambda) > 0$ for all $\lambda \in \Pi$.

5. Either the data $X_1, \ldots, X_T$ are untapered, i.e. $h_t = 1$ for all t in (2.1), or they are tapered with an asymptotically vanishing data taper of the form $h_t = h_\rho(\frac{t}{T})$ with $h_\rho(x) = u(x/\rho)\mathbf{1}_{[0,\rho/2]}(x) + \mathbf{1}_{[\rho/2,1-\rho/2]}(x) + u((1-x)/\rho)\mathbf{1}_{[1-\rho/2,1]}(x)$, where the function $u : [0,1] \to [0,1]$ is twice differentiable with bounded second derivative and $u(0) = 0$, $u(0.5) = 1$; $\rho = \rho(T)$ with $\rho(T) \sim T^{-\delta}$, where $\delta < 1/6$.

6. $\hat{f}$ is a spectral density estimator which converges to f uniformly, almost surely.

7. The innovations satisfy a Cramér condition: $\exists \delta > 0, d > 0$ such that for all $|t| > d$ the inequality $|E\exp(it\epsilon_1)| \leq 1 - \delta$ holds. Further, the limit of the dispersion matrix of $(H_{2,T}^{-1/2}(d_T(\omega_1), \ldots, d_T(\omega_k))^\top$ exists and is positive definite for fixed k. Furthermore, we assume $|a_v| \leq \gamma^{-|v|}$ for large $|v|$ and some $0 < \gamma < 1$.

8. $\hat{\kappa}_4$ is an estimator which converges almost surely to the fourth cumulant κ_4 of the process innovations ϵ_t.

The second part of Assumption 5 that the data taper is asymptotically vanishing is a realistic assumption. If an arbitrary taper is used the bootstrap procedure needs certain modifications. Assumption 7 is necessary for the proof of the convergence of the empirical distribution of the studentized periodogram values to the exponential distribution, which follows from Theorem 1 in Chen and Hannan (1980) for untapered data, or else from Theorem 4.3 in Janas and von Sachs (1995). It has to be made for the bootstrap method used in Section 5, but not for the goodness - of - fit tests in Sections 3 and 4, because here the resamples are drawn from an exponential distribution. It is not difficult to see that for e.g. a causal autoregressive model Assumptions 1-3 are fulfilled, provided the parameter set Θ satifies some restrictions.

To get an approximation of the test statistic V_T we first have to get an approximation of the Whittle estimate. By a Taylor expansion

$$\nabla_\theta \hat{\Delta}(\theta_0) = \nabla_\theta \hat{\Delta}(\theta_0) - \nabla_\theta \hat{\Delta}(\hat{\theta}) = \nabla_\theta^2 \hat{\Delta}(\overline{\theta})(\theta_0 - \hat{\theta}),$$

where $\overline{\theta}$ is between θ_0 and $\hat{\theta}$, we find that

$$\sqrt{T}(\hat{\theta}-\theta_0) = -\Gamma_0^{-1}\sqrt{T}\nabla_\theta\hat{\Delta}(\theta_0) + O_P(T^{-1/2}), \tag{2.2}$$

holds with $\Gamma_0 = \nabla_\theta^2\Delta(\theta_0)$. This follows from the convergence of $\hat{\theta}$ to θ_0 in probability (from $\hat{\Delta}(\theta) \to \Delta(\theta)$, uniformly for all $\theta \in \Theta$, cp. Dahlhaus, 1988), Assumption 3, and $\nabla_\theta^2\hat{\Delta}(\theta) = \nabla_\theta^2\Delta(\theta) + O_P(T^{-1/2})$ for all $\theta \in \Theta$.

To derive the distribution of the test statistic V_T we get (cp. Lemma 1 below) by another Taylor expansion for

$$V(\lambda) = \sqrt{T}(\frac{F_T(\lambda)}{F_T(\pi)} - \frac{F_{\hat{\theta}}(\lambda)}{F_{\hat{\theta}}(\pi)})$$

$$\begin{aligned} V(\lambda) &= \sqrt{T}(\frac{F_T(\lambda)}{F_T(\pi)} - \frac{F_{\theta_0}(\lambda)}{F_{\theta_0}(\pi)}) - \sqrt{T}(\hat{\theta}-\theta_0)^\top\nabla_\theta(\frac{F_{\theta_0}(\lambda)}{F_{\theta_0}(\pi)}) + O_P(T^{-1/2}) = \\ &= \sqrt{T}(\frac{F_T(\lambda)}{F_T(\pi)} - \frac{F_{\theta_0}(\lambda)}{F_{\theta_0}(\pi)}) + \sqrt{T}\nabla_\theta\hat{\Delta}(\theta_0)^\top\Gamma_0^{-1}\nabla_\theta(\frac{F_{\theta_0}(\lambda)}{F_{\theta_0}(\pi)}) + O_P(T^{-1/2}). \end{aligned} \tag{2.3}$$

Since

$$\nabla_\theta\hat{\Delta}(\theta_0) = \frac{1}{2\pi}\int_\Pi (I_T(\lambda) - f_{\theta_0}(\lambda))\nabla_\theta f_{\theta_0}^{-1},(\lambda)d\lambda$$

the second summand is of the form $c\sqrt{T}\int_\Pi(I_T(\lambda) - f_{\theta_0}(\lambda))\phi(\lambda)d\lambda$. It is well known that under H_0 this converges for a linear process to a normal distribution with mean zero and variance

$$2\pi c^2\int_\Pi \phi^2(\lambda)f_{\theta_0}^2(\lambda)d\lambda + c^2\frac{\kappa_4}{\sigma^4}(\int_\Pi \phi(\lambda)f_{\theta_0}(\lambda)d\lambda)^2,$$

where the second term is due to the (small but existing) dependence structure of the periodogram ordinates. Since this dependence structure is lost with an ordinary frequency domain bootstrap there is no chance of a good bootstrap approximation unless $\kappa_4 = 0$ or $\int_\Pi \phi(\lambda)f_{\theta_0}(\lambda)d\lambda = 0$ (cp. the discussion in Dahlhaus and Janas, 1996). Luckily the latter holds since it follows from Assumption 1 that $\int_\Pi f_{\theta_0}(\lambda)\nabla_\tau f_{\theta_0}^{-1}(\lambda)d\lambda = 0$ and $(\Gamma_0)_{\tau\sigma} = \nabla_\tau\nabla_\sigma\Delta(\theta_0) = 0$. The first term of (2.3) is a so called "ratio statistic" (cp. Dahlhaus and Janas, 1996) which can be approximated by a statistic of the form $c\sqrt{T}\int_\Pi(I_T(\lambda) - f_{\theta_0}(\lambda))\phi(\lambda)d\lambda$ with $\int_\Pi \phi(\lambda)f_{\theta_0}(\lambda)d\lambda = 0$. Heuristically this is the reason why the ordinary frequency bootstrap works for the statistic V_T. We make this precise in the following lemma.

Lemma 1 *Under Assumptions 1-5, the approximation*

$$V(\lambda) = \sqrt{T}(\frac{F_T(\lambda)}{F_T(\pi)} - \frac{F_{\hat{\theta}}(\lambda)}{F_{\hat{\theta}}(\pi)}) = \overline{V}(\lambda) + O_P(T^{-1/2})$$

holds uniformly for all $\lambda \in \Pi$, where

$$\begin{aligned}\overline{V}(\lambda) &= \sqrt{T}\int_\Pi \phi_\lambda(\alpha) I_T(\alpha)d\alpha + \sqrt{T}\int_\Pi \chi_{[0,\lambda]}(\alpha)(\frac{f(\alpha)}{\int_\Pi f(\gamma)d\gamma} - \frac{f_{\theta_0}(\alpha)}{\int_\Pi f_{\theta_0}(\gamma)d\gamma})d\alpha + \\ &\quad + \sqrt{T}(\frac{1}{2\pi}\int_\Pi (I_T(\alpha) - f_{\theta_0}(\alpha))\nabla_\theta f_{\theta_0}^{-1}(\alpha)d\alpha)^\top \Gamma_0^{-1}\nabla_\theta(\frac{F_{\theta_0}(\lambda)}{F_{\theta_0}(\pi)})\end{aligned}$$

and

$$\phi_\lambda(\alpha) = (\chi_{[0,\lambda]}(\alpha) - \frac{\int_\Pi \chi_{[0,\lambda]}(\gamma)f(\gamma)d\gamma}{\int_\Pi f(\gamma)d\gamma})/\int_\Pi f(\gamma)d\gamma.$$

Under the hypothesis H_0 we have

$$\overline{V}(\lambda) = \sqrt{T}\int_\Pi (\phi_\lambda(\alpha) + \omega_\lambda(\alpha))I_T(\alpha)d\alpha, \tag{2.4}$$

where

$$\omega_\lambda(\alpha) = \frac{1}{2\pi}(\nabla_\theta f_{\theta_0}^{-1}(\alpha))^\top \Gamma_0^{-1}\nabla_\theta(\frac{F_{\theta_0}(\lambda)}{F_{\theta_0}(\pi)}). \tag{2.5}$$

In this case we have $\int_\Pi \phi_\lambda(\alpha)f(\alpha)d\alpha = 0$ and $\int_\Pi \omega_\lambda(\alpha)f(\alpha)d\alpha = 0$.

Proof: see Appendix A.

We now derive the joint distribution of the approximations $\overline{V}(\lambda)$:

Lemma 2 *Under Assumptions 1-5, the approximations $\overline{V}(\lambda_1),\ldots,\overline{V}(\lambda_r)$ asymptotically have a multivariate normal distribution, and their cumulants are*

$$\begin{aligned}E\overline{V}(\lambda) &= \sqrt{T}\int_\Pi \chi_{[0,\lambda]}(\alpha)(\frac{f(\alpha)}{\int_\Pi f(\gamma)d\gamma} - \frac{f_{\theta_0}(\alpha)}{\int_\Pi f_{\theta_0}(\gamma)d\gamma})d\alpha + o(1), \\ Cov(\overline{V}(\lambda),\overline{V}(\mu)) &= 2\pi\int_\Pi (\phi_\lambda(\alpha)+\omega_\lambda(\alpha))(\phi_\mu(\alpha)+\omega_\mu(\alpha))f^2(\alpha)d\alpha + o(1), \\ Cum(\overline{V}(\lambda_1),\ldots,\overline{V}(\lambda_r)) &= o(1) \quad \textit{for } r > 2,\end{aligned}$$

uniformly for all $\lambda, \lambda_i, \mu \in \Pi$.

Proof: see Appendix B.

Under the hypothesis the expectation term vanishes asymptotically, so the distribution of $(\overline{V}(\lambda_1),\ldots,\overline{V}(\lambda_r))$ converges to a multivariate normal distribution with expectation zero. The expectation of $\overline{V}(\lambda)$ is diverging for some $\lambda \in \Pi$ under the alternative, so asymptotically it is possible to detect deviations from H_0.

Since $\overline{V}(\lambda)$ is an approximation of $V(\lambda)$ uniformly in λ, we have $V_T = \sup_{\lambda \in [0,\pi]} |\overline{V}(\lambda)| + O_P(T^{-1/2})$, so by the continuous mapping theorem the distribution of V_T converges under H_0 to the distribution of the supremum of the absolute value of a Gaussian process with zero mean and covariances as given in Lemma 2 (see Dahlhaus, 1988). This follows from the observation that the set $\{\overline{V}(\lambda), \lambda \in [0,\pi]\}$ can be interpreted as an empirical spectral process indexed by all functions $\{(\phi_\lambda(\cdot) + \omega_\lambda(\cdot)), \lambda \in \Pi\}$, which converges as a stochastic process to a Gaussian limit process under our assumptions (see Dahlhaus, 1988). The distribution of V_T is difficult to derive analytically, but it may be approximated by a bootstrap method as indicated in the next section.

3 Bootstrap of the test statistic

First we describe the frequency bootstrap which has been introduced in Franke and Härdle (1992); in Dahlhaus and Janas (1996) it is used for estimating the distribution of ratio statistics and Whittle estimates. The wild bootstrap, which will be described in Section 4, imitates the variance of the integrated periodogram correctly for quite arbitrary weight functions, even for linear processes with innovations that have non-vanishing fourth cumulant.

We get bootstrap resamples of the test statistic by a bootstrap of the periodogram ordinates. The bootstrap as suggested by Franke and Härdle (1992) is performed in the following way:

1. Calculate the periodogram values $I_T(\lambda_j)$ at the Fourier frequencies $\lambda_j = \frac{2\pi j}{T}$ for $j = 1, \ldots, n$, where $n = [T/2]$.
2. Calculate a uniformly consistent estimate $\hat{f}$ of the spectral density.
3. Compute the studentized periodogram values $\hat{\epsilon}_j = I_T(\lambda_j)/\hat{f}(\lambda_j)$ for $j = 1, \ldots, n$.
4. Rescale $\{\hat{\epsilon}_j\}$ and consider the approximately independent and identically distributed rescaled values $\{\tilde{\epsilon}_j\} = \{\hat{\epsilon}_j/\hat{\epsilon}.\}$, where $\hat{\epsilon}. = \frac{1}{n}\sum_{j=1}^{n} \hat{\epsilon}_j$.
5. Draw independent bootstrap replicates $\{\epsilon_j^*\}$ from the empirical distribution of the $\tilde{\epsilon}_j$.
6. Define the bootstrap periodogram values by $\{I_j^*\} = \{\hat{f}_j \cdot \epsilon_j^*\}$. Alternatively one may draw the $\{\epsilon_j^*\}$ from an exponential distribution with mean 1.

7. For a statistic $S(I)$ which is a function of the periodogram, estimate the distribution of $S(I) - S(f)$ by the empirical distribution of the bootstrap statistic $S(I^*) - S(\hat{f})$ after generating many independent periodogram resamples $I^* = (I_1^*, \ldots, I_n^*)$.

As we explain below Theorem 1 we recommend in this context to use the estimate $\hat{f} = f_{\hat{\theta}}$ and to draw the $\{\epsilon_j^*\}$ from an exponential distribution. This guarantees that the bootstrap test statistic has the correct distribution - even if the hypothesis is wrong. If we use this method then Assumption 7 becomes unnecessary.

An advantage of this bootstrap method over model-based bootstrap methods in the time domain is that it is not necessary to isolate innovations in the time domain which can be regarded as approximately iid, because this can be difficult for complex models. Instead one has to estimate the spectral density, which often is easier. Furthermore, one can use spectral densities for the hypothesis which are difficult to translate to the time domain. Of course this does not imply that a frequency domain bootstrap is generally preferable (cf. Chen and Romano, 1997).

For each resample I^* of the periodogram the Whittle estimate $\hat{\theta}^*$ has to be calculated by minimizing

$$\hat{\Delta}^*(\theta) = \frac{1}{2n} \sum_{j=1}^{n} (\log f_\theta(\lambda_j) + \frac{I^*(\lambda_j)}{f_\theta(\lambda_j)}),$$

over all $\theta \in \Theta$. $\hat{\Delta}^*(\theta)$ is interpreted as an estimate of the distance $\Delta^*(\theta)$ based on the estimated spectral density $\hat{f}$ instead of f,

$$\Delta^*(\theta) \quad = \quad \frac{1}{2\pi} \int_\Pi (\log f_\theta(\lambda) + \frac{\hat{f}(\lambda)}{f_\theta(\lambda)}) d\lambda$$

with minimizing value $\theta^* \in \Theta$. The statistical fluctuation of the periodogram I around the spectral density f is imitated by the fluctuation of I^* around the estimated spectral density $\hat{f}$.

We can estimate the distribution of V_T under the hypothesis by the bootstrap distribution (conditonal on the data) of the statistic V_T^* that is computed by substituting the periodogram in the definition of V_T by the bootstrap periodogram I^* and the Whittle estimator $\hat{\theta}$ by the Whittle estimator $\hat{\theta}^*$ calculated by minimizing $\hat{\Delta}^*$.

For every bootstrap resample and every Fourier frequency λ_j we can calculate

$$V^*(\lambda_j) = \sqrt{T} (\frac{F_T^*(\lambda_j)}{F_T^*(\pi)} - \frac{F_{\hat{\theta}^*}(\lambda_j)}{F_{\hat{\theta}^*}(\pi)}),$$

where the asterisk indicates that the resample is used for the calculation. Note that the bootstrap periodogram values are available only at the Fourier frequencies, so we have to replace integrals by sums: $F_T^*(\lambda_j) = \frac{2\pi}{T}\sum_{k=1}^{j} I_k^*$. We may also use such sums instead of integrals in the definition of $F_T(\lambda_j)$ so that its distribution can be expected to be imitated better by the bootstrap distribution of $F_T^*(\lambda_j)$, especially for small samples. It can be shown that the difference between sums and differences does not affect the asymptotic result of Theorem 1 in our case (cf. Dahlhaus, 1985, Section 3; Brillinger, 1981, Th. 5.10.2).

The distribution (conditional on the data) of the supremum V_T^* of the values $|V^*(\lambda_j)|$, $j = 1, \ldots, n$, is our estimate for the distribution of V_T under the hypothesis, because for $\hat{f} = f_{\theta^*}$ the approximation (2.4) of $V(\lambda)$ by $\overline{V}(\lambda)$ holds similarly for an approximation of $V^*(\lambda)$ by (under H_0)

$$\overline{V}^*(\lambda) = T^{1/2}\frac{2\pi}{T}\sum_{j=1}^{n}(\hat{\phi}_\lambda(\lambda_j) + \hat{\omega}_\lambda(\lambda_j))I_j^* \tag{3.1}$$

(see below). This can be shown as in the proof of Lemma 1 if I_T is substituted by I^*, f by $\hat{f}$, and θ_0 by θ^*.

It has been shown by Dahlhaus and Janas (1996) that the bootstrap method works for the Whittle estimator and for ratio statistics; with similar arguments in our case the convergence (in probability) under H_0 of the multivariate bootstrap distribution of $V^*(\lambda)$ to the multivariate distribution of the original values $V(\lambda)$ will follow for all finite sets of $\lambda \in \Pi$. Although the Fourier coefficients of the functions $(\phi_\lambda(\cdot)+\omega_\lambda(\cdot))$ do not fall exponentially as it is assumed in Dahlhaus and Janas (1996) (to ensure the validity of an Edgeworth expansion of the integrated periodogram), we still get the right covariances of the bootstrap resample, which means that the bootstrap "works":

Lemma 3 *Under Assumptions 1-6, the approximations $\overline{V}^*(\lambda)$ asymptotically have a normal distribution, and their cumulants are*

$$\begin{aligned}
E^*\overline{V}^*(\lambda) &= \sqrt{T}\int_\Pi \chi_{[0,\lambda]}(\alpha)\Big(\frac{\hat{f}(\alpha)}{\int_\Pi \hat{f}(\gamma)d\gamma} - \frac{f_{\theta^*}(\alpha)}{\int_\Pi f_{\theta^*}(\gamma)d\gamma}\Big)d\alpha + o(1), \quad a.s. \\
Cov^*(\overline{V}^*(\lambda), \overline{V}^*(\mu)) &= 2\pi\int_\Pi (\hat{\phi}_\lambda(\alpha) + \hat{\omega}_\lambda(\alpha))(\hat{\phi}_\mu(\alpha) + \hat{\omega}_\mu(\alpha))\hat{f}^2(\alpha)d\alpha + o(1) \to \\
&\to 2\pi\int_\Pi (\phi_\lambda(\alpha) + \omega_\lambda(\alpha))(\phi_\mu(\alpha) + \omega_\mu(\alpha))f^2(\alpha)d\alpha, \quad a.s. \\
Cum^*(\overline{V}^*(\lambda_1), \ldots, \overline{V}^*(\lambda_r)) &= o(1) \quad \text{for } r > 2, \quad a.s.
\end{aligned}$$

uniformly for all $\lambda, \lambda_i, \mu \in \Pi$. Here E^*, Cov^*, Cum^* denote the expectation, covariance and cumulants of the bootstrap statistics conditional on the data. $\hat{\phi}_\lambda$ and $\hat{\omega}_\lambda$ are defined as ϕ_λ and ω_λ are defined in Lemma 1 and (2.5), respectively, but with f replaced by $\hat{f}$ and θ_0 replaced by θ^*. We have $E^*\overline{V}^*(\lambda) \to 0$ for all $\lambda \in \Pi$ if $\hat{f} \in \{f_\theta, \theta \in \Theta\}$ (i.e. $\hat{f} = f_{\theta^*}$, which will be assumed in the following, cp. also the discussion below Theorem 1).

Proof: See Appendix C.

We can see from Lemma 3 that the bootstrap imitates the finite-dimensional distributions of the processes $\{\overline{V}(\lambda)\}$ and $\{V(\lambda)\}$, $\lambda \in \Pi$, and therefore the distribution of their absolute values is imitated, too, but we need to estimate the distribution of the supremum of them over all λ, which is our test statistic V_T. Theorem 1 states that the distribution of V_T is imitated by the bootstrap:

Theorem 1 *Suppose Assumptions 1-6 hold and the bootstrap is applied with $\hat{f} = f_{\hat{\theta}}$. Then, under H_0, the conditional distribution of the bootstrap statistic V_T^* converges to the distribution of the original statistic V_T a.s.:*

$$\sup_{x \in \mathbf{R}} |P(V_T \leq x) - P^*(V_T^* \leq x)| = o(1),$$

where P^* denotes the conditional probability. Therefore the test which rejects H_0 if V_T is larger than the $(1-\alpha)$-quantile of the conditional distribution of V_T^* asymptotically has the level α.

Proof: See Appendix D.

An important issue is the selection of the spectral density estimate $\hat{f}$. For several applications a nonparametric estimate is the right choice since this leads to a model - free bootstrap (cf. Dahlhaus and Janas, 1996). However, in the present context the situation is different: With our bootstrap test statistic V_T^* we want to estimate the distribution of the statistic V_T under the hypothesis H_0 - even if H_0 is wrong, i.e. even if V_T itself has a different distribution. This guarantees that the test has a good power. For this reason we recommend using $\hat{f} = f_{\hat{\theta}}$ and drawing the ϵ_j^* independently from an exponential distribution with mean 1 (which is the asymptotic distribution of $I_T(\lambda)/f(\lambda)$).

If instead we used a nonparametric estimate $\hat{f}$ then $\hat{f}(\cdot)/\int \hat{f}$ were close to $f_{\theta^*}(\cdot)/\int f_{\theta^*}$ also under the alternative leading to a low power of the test. Using exponentially distributed

ϵ_j^* avoids the distortion of the residuals which will result if an inconsistent spectral density estimate is used (in particular under the alternative).

There is a second reason for using $\hat{f} = f_{\hat{\theta}}$ instead of a nonparametric estimator. Under the hypothesis H_0 we have $f = f_{\theta_0}$ (in the "real world") - but $\hat{f} = f_{\theta^*}$ (in the "bootstrap world") if and only if $\hat{f}$ is a parametric estimate itself. This implies that the conditional expectation $E^*\overline{V}^*(\lambda)$ converges to zero under H_0 which is necessary for the bootstrap to work (see Lemma 3).

Dahlhaus and Janas (1996) prove that the bootstrap approximation of statistics of the form $\int_\Pi \psi(\alpha) I_T(\alpha) d\alpha$ with $\int_\Pi \psi(\alpha) f(\alpha) d\alpha = 0$ and $\psi(\alpha)$ smooth even leads to a better approximation than a normal approximation. We conjecture that this also holds for the approximation of the distribution of V_T. However, in the present situation the benefit of the bootstrap is even greater since the variance of the asymptotic distribution of V_T is too difficult to evaluate.

Simulation results, which give an impression of the performance of this modified test, and an example can be found in Section 6.

4 Wild bootstrap tests

Since the estimated variance of the innovations cancels out in the above test statistic the test has no power for wrong σ^2. Only the other parameters (i.e. the shape of the spectral density) can be tested. To overcome this restriction we discuss in the present chapter a bootstrap approximation of the test statistic

$$W_T = \sup_{\lambda \in [0,\pi]} |\sqrt{T}(F_T(\lambda) - F_{\hat{\theta}}(\lambda))|.$$

The hypothesis then is

$$H_0 : \theta \in \Theta_0 = \{\theta | \theta = (\sigma^2, \tau_1, \ldots, \tau_p)^\top, \sigma^2 \in [a,b]\} \subset \Theta, b \geq a,$$

i.e. the test should reject if the variance σ^2 is not between a and b. To this purpose the bootstrap method must imitate the distribution of $\hat{\sigma}^2$, which cannot be done by the above bootstrap except in special cases (e.g. for Gaussian processes). The reason is that the part of the variance of the integrated periodogram which stems from the correlation of different periodogram ordinates (and which depends on the fourth cumulant of the innovations) is ignored by independent resampling.

For this purpose Janas and Dahlhaus (1994) use the wild bootstrap which "artificially" introduces a dependence between the different resampled periodogram values: Instead of generating the periodogram resample by $I_j^* := \hat{f}_j \cdot \epsilon_j^*$, as in Section 3, now a consistent estimate $\hat{\kappa}_4$ of the fourth cumulant κ_4 of the process innovations ϵ_t is calculated (cf. Grenander and Rosenblatt, 1956, Ch.5.6), and the resamples are generated by

$$I_j^* \quad := \quad \hat{f}_j(\epsilon_j^* + c^*), \tag{4.1}$$

where

$$c^* \quad = \quad ((1 + 0.5(\hat{\kappa}_4/\hat{\sigma}^4))^{1/2} - 1)\frac{1}{n}\sum_{k=1}^{n}(\epsilon_k^* - 1). \tag{4.2}$$

c^* introduces a small correction to the bootstrap - periodogram which emulates the correlation structure of the true periodogram. This provides an additional variance for the integrated periodogram which stems from the correlation between different periodogram ordinates; $\hat{\sigma}^2$ is the estimated variance of the innovations, which converges to σ^2 under H_0 (see Assumption 4).

The resulting bootstrap periodogram values have (conditional) covariances converging to

$$Cov^*(I_j^*, I_k^*) = f_j f_k(\delta_{j,k} + \kappa_4/T)$$

if $\hat{f} \to f$, $\hat{\sigma}^2 \to \sigma^2$, and $\hat{\kappa}_4 \to \kappa_4$. Note that c^* in (4.1) depends on the resample values $\{\epsilon_k^*\}$, but it is independent of the index j. Janas and Dahlhaus (1994), Section 3, show that with this bootstrap method a convergence in the Mallows d_2-metric of the original statistics and the bootstrap statistics is achieved (almost surely) for statistics which are integrated periodograms $\int_\Pi \phi(\lambda) I_T(\lambda) d\lambda$; this implies that their distribution functions converge to the same limit.

The distribution of the test statistic W_T now is estimated by the conditional distribution of $W_T^* = \sup_{j=1,\ldots,n} |T^{1/2}(F_T^*(\lambda_j) - F_{\hat{\theta}^*}(\lambda_j))|$, which is calculated as in Section 3 from the bootstrap resamples I_j^* as in (4.1).

Of course we require that $\hat{\theta}, \hat{\theta}^* \in \Theta_0$, which means in particular that $\hat{\sigma}^2$ must lie in $[a, b]$. Under the hypothesis, $H_0 : f \in \Theta_0$, the estimate $\hat{\sigma}^2$ will converge to $\sigma_0^2 \in [a, b]$. Using

$\int_\Pi f(\lambda)/f_{\theta_0}(\lambda)d\lambda = \pi$, $\frac{\partial}{\partial\sigma^2} f_{\theta_0}^{-1}(\lambda) = -\frac{1}{\sigma_0^2} f_{\theta_0}^{-1}(\lambda)$ (see Assumption 1), and (2.2), we get the following approximation, similarly as in Appendix A:

$$\begin{aligned}
W(\lambda) &:= \sqrt{T}(F_T(\lambda) - F_{\hat\theta}(\lambda)) = \qquad (4.3)\\
&= \sqrt{T}(\int_\Pi \chi_{[0,\lambda]}(\alpha)(I_T(\alpha) - f_{\theta_0}(\alpha))d\alpha) - \sqrt{T}((\nabla_\theta F_{\theta_0}(\lambda))^\top(\hat\theta - \theta_0) + O_P(T^{-1/2}) = \\
&= \sqrt{T}(\int_\Pi \chi_{[0,\lambda]}(\alpha)(I_T(\alpha) - f(\alpha))d\alpha) + \sqrt{T}(\int_\Pi \chi_{[0,\lambda]}(\alpha)(f(\alpha) - f_{\theta_0}(\alpha))d\alpha) + \\
&\quad + \sqrt{T}(\nabla_\theta F_{\theta_0}(\lambda))^\top \Gamma_0^{-1} \frac{1}{2\pi}(\frac{-1}{\sigma_0^2}\int_\Pi f_{\theta_0}^{-1}(\alpha)(I_T(\alpha) - f(\alpha))d\alpha, \\
&\qquad \int_\Pi (\nabla_\tau f_{\theta_0}^{-1}(\alpha))^\top (I_T(\alpha) - f(\alpha))d\alpha)^\top + O_P(T^{-1/2}) = \\
&= \overline{W}(\lambda) + O_P(T^{-1/2}).
\end{aligned}$$

Here we use $\Gamma_0 = \nabla_\theta^2 \Delta(\theta_0)$ and

$$\nabla_\theta \hat\Delta(\theta_0) = \frac{1}{2\pi}(\frac{-1}{\sigma_0^2}\int_\Pi f_{\theta_0}^{-1}(\alpha)(I_T(\alpha) - f(\alpha))d\alpha, \int_\Pi (\nabla_\tau f_{\theta_0}^{-1}(\alpha))^\top (I_T(\alpha) - f(\alpha))d\alpha)^\top.$$

The error term is uniformly of order $O_P(T^{-1/2})$ for all $\lambda \in \Pi$.

The above derivation assumes that $\sigma_0^2 \in [a,b]$, which is true under H_0. Otherwise, we will have no convergence $\hat\sigma^2 \to \sigma_0^2$, so the Taylor expansion does not work. Under our Assumptions 1-8, the wild bootstrap will imitate the multivariate distribution of the statistics $\overline{W}(\lambda_j)$ for finite sets $\{\lambda_1, \ldots, \lambda_r\}$ under the hypothesis (see Janas and Dahlhaus, 1994).

Lemma 4 *Under Assumptions 1-6 and 8, and $\sigma_0^2 \in [a,b]$, the approximations $\overline{W}(\lambda)$ asymptotically have a normal distribution, and their cumulants are*

$$\begin{aligned}
E\overline{W}(\lambda) &= \sqrt{T}\int_\Pi \chi_{[0,\lambda]}(\alpha)(f(\alpha) - f_{\theta_0}(\alpha))d\alpha + o(1), \\
Cov(\overline{W}(\lambda), \overline{W}(\mu)) &= 2\pi \int_\Pi f^2(\alpha)\psi_\lambda(\alpha)\psi_\mu(\alpha)d\alpha + \\
&\quad + \frac{\kappa_4}{\sigma^4}(\int_\Pi f(\alpha)\psi_\lambda(\alpha)d\alpha)(\int_\Pi f(\alpha)\psi_\mu(\alpha)d\alpha) + o(1), \\
Cum(\overline{W}(\lambda_1), \ldots, \overline{W}(\lambda_r)) &= o(1) \quad \text{for } r > 2, \quad \text{where} \\
\psi_\lambda(\alpha) &:= \chi_{[0,\lambda]}(\alpha) + (\nabla_\theta F_{\theta_0}(\lambda))^\top \Gamma_0^{-1}\frac{1}{2\pi}(-\frac{1}{\sigma_0^2} f_{\theta_0}^{-1}(\alpha), (\nabla_\tau f_{\theta_0}^{-1}(\alpha))^\top)^\top.
\end{aligned}$$

The remainder terms converge uniformly for all $\lambda, \lambda_i, \mu \in \Pi$.

Proof: see Appendix B.

In the covariance expression the fourth cumulant κ_4 appears, which shows that the ordinary frequency bootstrap will not work without a correction as in (4.1), because $\kappa_4 \int \psi_\lambda f$

cannot be expected to vanish in general. The cumulants of the bootstrap statistics $\overline{W}^*(\lambda)$, which are defined analogously to $\overline{V}^*(\lambda)$ in Section 3, converge to the same limits a.s. if a spectral density estimate is used which is consistent under the hypothesis. As in Section 3 one should use a parametric spectral density estimate $\hat{f} = f_{\hat{\theta}}$, $\hat{\theta} \in \Theta_0$, for the bootstrap.

We replace f, θ_0, I_T, $\hat{\theta}$ by $\hat{f}$, θ^*, I^*, $\hat{\theta}^*$, respectively, and use a Taylor expansion as in (4.3) to calculate the approximations $\overline{W}^*(\lambda_j)$. The properties of these approximations under the hypothesis are summarized in the following lemma.

Lemma 5 *Under Assumptions 1-6 and 8, and H_0, the approximations*

$$\overline{W}^*(\lambda) = \sqrt{T}(F_T^*(\lambda) - F_{\hat{\theta}^*}(\lambda))$$

asymptotically have a normal distribution, and their cumulants are

$$\begin{aligned}
E^*\overline{W}^*(\lambda) &= \sqrt{T}\int_\Pi \chi_{[0,\lambda]}(\alpha)(\hat{f}(\alpha) - f_{\theta^*}(\alpha))d\alpha + o(1), \\
Cov^*(\overline{W}^*(\lambda), \overline{W}^*(\mu)) &= 2\pi(\int_\Pi \hat{\psi}_\lambda(\alpha)\hat{\psi}_\mu(\alpha)\hat{f}^2(\alpha)d\alpha + \\
& \quad +(\hat{\kappa}_4/\hat{\sigma}^4)(\int_\Pi \hat{f}(\alpha)\hat{\psi}_\lambda(\alpha)d\alpha)(\int_\Pi \hat{f}(\alpha)\hat{\psi}_\mu(\alpha)d\alpha) + o(1) \to \\
&\to 2\pi(\int_\Pi \psi_\lambda(\alpha)\psi_\mu(\alpha)f^2(\alpha)d\alpha + \\
& \quad +(\kappa_4/\sigma^4)(\int_\Pi f(\alpha)\psi_\lambda(\alpha)d\alpha)(\int_\Pi f(\alpha)\psi_\mu(\alpha)d\alpha), \\
Cum^*(\overline{W}^*(\lambda_1), \ldots, \overline{W}^*(\lambda_r)) &= o(1) \quad \text{for } r > 2,
\end{aligned}$$

uniformly for all $\lambda, \lambda_i, \mu \in \Pi$ (always a.s.). $\hat{\psi}_\lambda$ is defined as ψ_λ is defined in Lemma 4, but with f replaced by $\hat{f}$ and θ_0 replaced by θ^*. We have $E^*\overline{W}^*(\lambda) \to 0$ if $\hat{f} = f_{\theta^*}$.

Proof: See Appendix C.

The equicontinuity of the bootstrap statistics $\overline{W}^*(\lambda)$, which is necessary to get a valid approximation of the supremum W_T^* of the absolute values of these bootstrap statistics, follows similarly as for V_T^* in Section 3.

Theorem 2 *Suppose Assumptions 1-6 and 8 hold, and the wild bootstrap is applied with $\hat{f} = f_{\hat{\theta}}$. Then, under H_0, the conditional distribution of the bootstrap statistic W_T^* converges to the distribution of the original statistic W_T a.s.:*

$$\sup_{x \in \mathbf{R}} |P(W_T \le x) - P^*(W_T^* \le x)| = o(1),$$

where P^* denotes the conditional probability. Therefore the test which rejects H_0 if W_T is larger than the $(1-\alpha)$-quantile of the conditional distribution of W_T^* asymptotically has the level α.

Proof: See Appendix D.

Of course, one can also test the hypothesis that $f = f_0$, with some given spectral density f_0, and the distribution of the test statistic $\sup_{\lambda\in\Pi} |\sqrt{T}(F_T(\lambda) - \int_\Pi \chi_{[0,\lambda]}(\alpha) f_0(\alpha) d\alpha)|$ is imitated by the wild bootstrap (with $\hat{f} = f_0$) under H_0, too.

5 Comparison of two time series

Now suppose we have two independent time series $X_1^{(1)}, \ldots, X_{T_1}^{(1)}$ and $X_1^{(2)}, \ldots, X_{T_2}^{(2)}$, where T_1 and T_2 may be different, and we want to test whether their spectral densities are equal,

$$H_0 : f^{(1)}(\lambda) = f^{(2)}(\lambda) \text{ for all } \lambda \in \Pi,$$

where $f^{(i)}$ is the spectral density of time series i, or we want to test whether the shapes of their spectral densities are equal,

$$H_0^F : f^{(1)}(\lambda)/f^{(2)}(\lambda) = \text{const. for all } \lambda \in \Pi,$$

against the alternatives $f^{(1)}(\lambda) \neq f^{(2)}(\lambda)$ and $f^{(1)}(\lambda)/f^{(2)}(\lambda) \neq f^{(1)}(\mu)/f^{(2)}(\mu)$ for some $\lambda, \mu \in \Pi$, respectively.

For tests of the hypothesis H_0 Diggle and Fisher (1991) use the empirical spectral process after dividing by the integrated periodograms. Their test statistic (we again focus on tests of Kolmogorov-Smirnov type) is essentially

$$D = \sup_{\lambda\in\Pi} \left| \frac{F_{T_1}^{(1)}(\lambda)}{F_{T_1}^{(1)}(\pi)} - \frac{F_{T_2}^{(2)}(\lambda)}{F_{T_2}^{(2)}(\pi)} \right|,$$

where $F_T^{(i)}(\lambda) = \int_\Pi \chi_{[0,\lambda]}(\alpha) I_T^{(i)}(\alpha) d\alpha$. Diggle and Fisher estimate its distribution by exchanging the periodogram values $I_j^{(1)}$, $I_j^{(2)}$ of the two time series for every j with probability 0.5. Under H_0 the periodograms are independent and (approximately) identically distributed. Of course, this does not work under H_0^F since the periodogram values may have different distributions. Further, the test does not work without modifications for time series with unequal numbers of observations. Due to the division by the integrated periodogram, the test has no power against a difference in variance. Simply omitting the division will not work in general, because then the dependence between different periodogram ordinates of one time series no longer can be neglected as in the last section. In this case the method of Diggle and Fisher (1991) will produce a critical value different from the true one, even asymptotically. To overcome these restrictions we propose the following bootstrap methods.

5.1 Test of H_0^F

For a test which is insensitive against a different scale of the spectral densities we use the above test statistic D, but we propose a bootstrap method to estimate the distribution of D under the hypothesis.

The distribution of D is estimated under H_0^F by the bootstrap distribution of

$$\begin{aligned} D^* &= \sup_\lambda \left| \frac{F_{T_1}^{(1)*}(\lambda)}{F_{T_1}^{(1)*}(\pi)} - \frac{F_{T_2}^{(2)*}(\lambda)}{F_{T_2}^{(2)*}(\pi)} \right|, \quad \text{where} \\ F_T^{(i)*}(\lambda) &= \frac{1}{n_i} \sum_{j=1}^{n_i} \chi_{[0,\lambda]}(\lambda_j) I_j^{(i)*}, \\ n_i &= [T_i/2]. \end{aligned}$$

For the generation of the bootstrap periodogram values $I_j^{(i)*}$ we suggest the following method: First compute a nonparametric (consistent) spectral density estimate $\hat{f}^{(i)}$ for each of the two time series, $i = 1, 2$, and calculate the residuals $\hat{\epsilon}_j^{(i)}$, $j = 1, \ldots, n_i$, as indicated in Section 3, separately for each time series. Under H_0^F, $f^{(1)}(\lambda)/\int_\Pi f^{(1)}(\gamma)d\gamma = f^{(2)}(\lambda)/\int_\Pi f^{(2)}(\gamma)d\gamma$ holds, and therefore

$$g^{(i)}(\lambda) = \left[\frac{T_1}{T_1+T_2}\frac{f^{(1)}(\lambda)}{\int_\Pi f^{(1)}(\gamma)d\gamma} + \frac{T_2}{T_1+T_2}\frac{f^{(2)}(\lambda)}{\int_\Pi f^{(2)}(\gamma)d\gamma}\right]\int_\Pi f^{(i)}(\gamma)d\gamma$$

equals $f^{(i)}(\lambda)$ for all λ. Calculate the corresponding estimators $\hat{g}^{(i)}(\lambda)$ by replacing $f^{(1)}$ and $f^{(2)}$ by $\hat{f}^{(1)}$ and $\hat{f}^{(2)}$, respectively, and generate the resamples $I_j^{(i)*} = \hat{g}^{(i)}(\lambda_j) \cdot \epsilon_j^{(i)*}$, $j = 1, \ldots, n_i$, where $\epsilon_j^{(i)*}$ are resampled independently from $\tilde{\epsilon}_1^{(i)}, \ldots, \tilde{\epsilon}_{n_i}^{(i)}$. We have a bootstrap method which assumes that H_0^F is true, because $\hat{g}^{(1)}(\lambda)/(\int \hat{g}^{(1)}(\gamma) = \hat{g}^{(2)}(\lambda)/\int \hat{g}^{(2)}(\gamma))$ holds. This procedure guarantees that D^* has asymptotically the same distribution as D has under H_0^F - even if the hypothesis is not true. We prove this assertion under H_0^F in the following theorem.

Theorem 3 *Under H_0^F and with Assumptions 1-7 holding for each of the two independent time series $X_1^{(1)}, \ldots, X_{T_1}^{(1)}$ and $X_1^{(2)}, \ldots, X_{T_2}^{(2)}$, the conditional distribution of the bootstrap statistic $(\frac{T_1 T_2}{T_1+T_2})^{1/2} D^*$ converges to the same limit as the original statistic $(\frac{T_1 T_2}{T_1+T_2})^{1/2} D$ a.s. if $\min(T_1, T_2) \to \infty$ holds. Therefore the test which rejects H_0^F if D is larger than the $(1-\alpha)$-quantile of the conditional distribution of D^* asymptotically has the level α.*

Proof: See Appendix E.

5.2 Test of H_0

If the variances of the processes are included in the hypothesis, we can use the obvious modification of D,

$$E = \sup_{\lambda \in \Pi} |F_{T_1}^{(1)}(\lambda) - F_{T_2}^{(2)}(\lambda)|.$$

As we have seen in Section 4, for statistics of this type we have to use the wild bootstrap to get the correct variance, except if the fourth cumulants $\kappa_4^{(1)}$, $\kappa_4^{(2)}$ of the innovations of both processes vanish.

The resampling of the periodogram ordinates works similarly as in the test of H_0^F, but now we can use

$$g(\lambda) := \frac{T_1}{T_1+T_2} f^{(1)}(\lambda) + \frac{T_2}{T_1+T_2} f^{(2)}(\lambda),$$

(and correspondingly $\hat{g}(\lambda) := \frac{T_1}{T_1+T_2} \hat{f}^{(1)}(\lambda) + \frac{T_2}{T_1+T_2} \hat{f}^{(2)}(\lambda)$) instead of $g^{(i)}(\lambda)$, which equals both $f^{(1)}(\lambda)$ and $f^{(2)}(\lambda)$ under H_0. When resampling $I_j^{(i)*}$, we have to add a term containing $\hat{\kappa}_4^{(i)}$ which corrects the variance of the integrated periodogram (see Section 4); if we include $\kappa_4^{(1)} = \kappa_4^{(2)}$ in the hypothesis H_0, we can estimate $\kappa_4^{(i)}$ from both time series, otherwise we have to estimate it separately for each time series.

Defining $E^* = \sup_{\lambda \in \Pi} |F_{T_1}^{(1)*}(\lambda) - F_{T_2}^{(2)*}(\lambda)|$ with this bootstrap, we get the following theorem:

Theorem 4 *Under H_0 and with Assumptions 1-8 holding for each of the two independent time series $X_1^{(1)}, \ldots, X_{T_1}^{(1)}$ and $X_1^{(2)}, \ldots, X_{T_2}^{(2)}$, the conditional distribution of the bootstrap statistic $(\frac{T_1 T_2}{T_1+cT_2})^{1/2} E^*$ converges to the same limit as the the original statistic $(\frac{T_1 T_2}{T_1+cT_2})^{1/2} E$ a.s. if $\min(T_1, T_2) \to \infty$ holds; $c = c(\kappa_4^{(1)}, \kappa_4^{(2)})$ is some positive constant. Therefore the test which rejects H_0 if E is larger than the $(1-\alpha)$-quantile of the conditional distribution of E^* asymptotically has the level α.*

Proof: See Appendix E.

It should be noted that in practice D and E are calculated as suprema over all Fourier frequencies $\lambda_j^{(1)}$, $\lambda_k^{(2)}$, $j = 1, \ldots, n_1$, $k = 1, \ldots, n_2$, of the two time series instead over all $\lambda \in \Pi$, at least for the resamples, because only at these points the bootstrap periodogram values are available. If $T_1 \neq T_2$, the Fourier frequencies of the time series lie at different locations; in this case the statistics may be calculated as suprema of the absolute values of e.g. the linearly interpolated normalized periodogram values.

6 Simulations and examples

In this section the performance of the goodness-of-fit test of Theorem 1 and of the comparison test of Theorem 3 are illustrated for small samples by a simulation study. We do not include simulations for the tests involving the wild bootstrap.

For small samples several additional considerations are necessary to improve the performance of the bootstrap. These concern the selection of a suitable nonparametric spectral density estimate (for comparison tests) and some small sample corrections. The latter e.g. try to take into account deviations from the asymptotic distribution that stem from the data taper or the bias of the residuals originating from the use of the estimated spectral density instead of the unknown true spectral density. For these and other small sample corrections see Section 6.3.

6.1 Simulations

6.1.1 Goodness-of-fit tests

First we want to explore the small sample performance of the goodness-of-fit test described in Sections 2 and 3 by a small simulation study. For this purpose for each pair of several AR processes and one ARMA process, and of several AR model classes, 2000 samples have been simulated, each sample with 128 observations. The innovations of all simulated processes are uniformly distributed with mean 0 and variance 1; note that these random variables have fourth cumulants $-1.2 \neq 0$. For every sample 1000 resamples are generated using the bootstrap method described in Section 3, based on a parametric spectral density estimate within the model class under consideration (estimated by the Yule-Walker estimator), and these resamples are used to estimate the distribution of the test statistic under the hypothesis; as hypotheses we use AR(0),...,AR(6) model classes. The inverse roots of the characteristic polynomials $1 - \sum_{j=1}^{p} \theta_j z^j$ of the AR parts of order p, $p = 1, \ldots, 6$, of these models have modulus 0.9 and the following phases:

$p = 1$: 0;

$p = 2$: $\pi/2$, $-\pi/2$;

$p = 3$: 0, $\pi/2$, $-\pi/2$;

$p = 4$: $\pi/4$, $-\pi/4$, $3\pi/4$, $-3\pi/4$;

$p = 5$: 0, $\pi/4$, $-\pi/4$, $3\pi/4$, $-3\pi/4$;

$p = 6$: $\pi/2$, $-\pi/2$, $\pi/4$, $-\pi/4$, $3\pi/4$, $-3\pi/4$.

The normalized spectral densities of these models are plotted in Figure 1. The modulus of the roots of the AR(2) model has been varied further (0.5, 0.7) to get some impression about the effect of spectral peaks of different heights on the bootstrap test. Further, an ARMA(2,1) model has been included as an example of a process for which all model classes under investigation are misspecified. The power of the test is estimated by the relative frequency of rejections for the 2000 simulated samples, and the critical values of the test are determined separately for every sample by the bootstrap. The levels of significance have been chosen as 2.5%, 5%, and 10%. In Table 1 one can see that the tests generally perform reasonably well. There usually is a good agreement between the nominal and the estimated level of significance, if one considers that the standard deviation of the power is about 0.35%, 0.49%, and 0.67% for the true models and the different levels; the tests for model classes with several parameters tend to be too conservative, though. It is not possible to identify models with too many parameters; for this purpose one has to use a model selection method (such as using a model selection criterium like AIC etc). As these models are too large, but not really "wrong", a goodness-of-fit test does not have to reject in these cases.

Potential problems include the deviation of the true distribution of the periodogram values from the asymptotic exponential distribution, correlation between the different periodogram values, and the bias of the spectral density estimate that is the base of the bootstrap procedure. These difficulties will arise especially when the sample size is small. The use of a good parametric spectral estimate, e.g. a Yule-Walker estimate with data taper (used here) or a Burg estimator, which have a smaller bias than the Yule-Walker estimator with no data taper, is therefore recommended. A bootstrap calibration procedure (see Section 6.3) or a bias correction of the spectral density estimate may further reduce these small sample problems, but have not been used in these simulations.

Chen and Romano (1997) report that smoothing the periodogram or testing the residuals of the fitted model for white noise increases the power of their test method; this may also be true for our method.

6.1.2 Comparison tests

Now we use the AR(i) models to illustrate the performance of the comparison test of Theorem 3. For every pair (i, j), $i, j = 0, \ldots, 6$, we generate 5000 samples, each consisting of one realisation of the AR(i) process of length 128 and one realisation of the AR(j) process (independent of the other process), also of length 128, with parameters as given in Table 1. The variance of both processes' innovations is 1, but using different variances does not change the result as only the shapes of the spectral densities (cp. Figure 1) are compared. For every sample we calculate the test statistic D, and we estimate its distribution under the hypothesis H_0^F by drawing 1000 resamples as indicated in Section 5.1, separately for each sample. Then we estimate the power of the test (at level 5%) by the number of rejections of the hypothesis divided by 5000; the results are given in Table 2 for all pairs (i, j) of models. Note that the standard deviation of the estimated power is 0.3% under the hypothesis. The simulation is repeated for time series of lengths 128 and 256, respectively (see Table 3).

One can see that the level of the tests are reasonably close to the nominal level of 5%. However, it seems to be difficult to reject the hypothesis for some pairs of models, e.g. the AR(2) and the AR(6), or the AR(0) and the AR(4) models. This can be seen from the low power of the tests for these pairs in Tables 2 and 3 for the comparison tests, and in Table 1 for the goodness-of-fit tests (with the higher order model as true model). The spectral densities of these models are not different enough to be discriminated by the test for a moderate sample size (cf. Figure 1).

6.2 Real data example

As illustration we use the Beveridge Wheat Prize Index, which is an annual index of European wheat prizes for the years 1500-1869. After trend correction (cp. Figure 2 and Anderson, 1971, Appendix A.1) and subtraction of the mean we analyze these data by goodness-of-fit tests. Furthermore, we compare the subset of the data for the years 1500-1599 with the

level= 2.5%	fitted models						
true models	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
AR(0)	2.8	3.7	1.6	1.7	1.4	1.0	0.7
AR(1)	100.0	4.1	4.2	4.4	4.4	4.7	4.6
AR(2,0.5)	2.4	7.5	2.4	2.0	0.8	0.7	0.4
AR(2,0.7)	13.5	50.0	2.0	0.8	0.8	0.6	0.4
AR(2)	100.0	100.0	2.0	1.7	1.7	1.2	1.4
AR(3)	100.0	76.5	99.2	1.6	2.0	1.8	1.9
AR(4)	60.0	96.0	99.5	99.4	0.7	0.8	0.7
AR(5)	100.0	22.8	20.2	34.0	70.6	1.8	1.7
AR(6)	92.9	100.0	10.3	17.3	55.3	58.9	0.6
ARMA(2,1)	100.0	100.0	28.8	2.7	3.4	2.0	1.0

level= 5%	fitted models						
true models	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
AR(0)	5.7	6.1	3.8	4.0	2.9	2.2	1.6
AR(1)	100.0	7.7	7.7	7.7	7.4	7.7	7.6
AR(2,0.5)	5.9	13.1	4.9	3.4	2.1	1.4	1.5
AR(2,0.7)	23.4	62.7	4.1	2.9	2.2	2.1	1.6
AR(2)	100.0	100.0	4.3	3.5	4.1	3.3	3.2
AR(3)	100.0	87.5	99.7	4.0	4.8	4.1	3.4
AR(4)	72.2	98.5	100.0	99.9	2.0	2.1	1.6
AR(5)	100.0	32.1	30.6	46.0	80.9	3.9	3.7
AR(6)	99.5	100.0	18.5	26.9	73.3	76.1	1.9
ARMA(2,1)	100.0	100.0	39.6	7.2	6.0	4.1	2.8

level= 10%	fitted models						
true models	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
AR(0)	11.0	12.1	9.7	8.5	6.1	5.4	4.3
AR(1)	100.0	13.6	13.3	13.6	13.9	13.4	13.0
AR(2,0.5)	12.7	21.8	8.5	7.1	5.0	3.6	3.1
AR(2,0.7)	39.8	91.0	9.2	6.8	5.8	4.9	4.3
AR(2)	100.0	100.0	8.2	9.3	9.8	7.7	7.7
AR(3)	100.0	94.1	99.8	9.3	9.6	8.5	8.6
AR(4)	86.3	99.7	100.0	100.0	5.3	5.2	5.2
AR(5)	100.0	45.7	44.2	58.4	89.3	8.5	8.1
AR(6)	100.0	100.0	32.1	41.7	86.7	88.4	4.7
ARMA(2,1)	100.0	100.0	54.0	13.7	11.5	8.0	7.4

Table 1: Estimated power (in percent) of goodness-of-fit tests for various models (true model vertical, fitted model horizontal) and levels 2.5%, 5%, and 10%; 2000 samples with 128 observations, 1000 resamples for each sample.

The true models are (ϵ_t iid uniformly distributed with mean 0 and variance 1):

AR(0): $X_t = \epsilon_t$,

AR(1): $X_t = 0.9X_{t-1} + \epsilon_t$,

AR(2): $X_t = -0.81X_{t-2} + \epsilon_t$,

AR(3): $X_t = 0.9X_{t-1} - 0.8X_{t-2} + 0.72X_{t-3} + \epsilon_t$,

AR(4): $X_t = -0.6561X_{t-4} + \epsilon_t$,

AR(5): $X_t = 0.9X_{t-1} - 0.6561X_{t-4} + 0.59049X_{t-5} + \epsilon_t$,

AR(6): $X_t = -0.81X_{t-2} - 0.6561X_{t-4} - 0.531441X_{t-6} + \epsilon_t$,

ARMA$(2,1)$: $X_t = -0.81X_{t-2} + \epsilon_t + 0.5\epsilon_{t-1}$,

AR$(2,h)$: $X_t = -h^2 \cdot X_{t-2} + \epsilon_t$, $h = 0.5, 0.7$.

level=5%	128 data						
128 data	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
AR(0)	5.7	100.0	97.2	97.8	27.3	100.0	45.2
AR(1)	-	4.5	100.0	52.3	100.0	38.5	100.0
AR(2)	-	-	4.4	87.9	100.0	100.0	19.1
AR(3)	-	-	-	5.1	99.3	57.2	91.1
AR(4)	-	-	-	-	5.4	99.7	62.3
AR(5)	-	-	-	-	-	5.9	100.0
AR(6)	-	-	-	-	-	-	4.3

Table 2: Estimated power of comparison tests for AR models (in percent). For every model 5000 samples with 128 observations have been calculated, and 1000 resamples have been used for each sample.

level=5%	256 data						
128 data	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
AR(0)	5.3	100.0	97.7	99.4	27.2	100.0	53.9
AR(1)	100.0	5.6	100.0	35.7	100.0	24.2	100.0
AR(2)	98.7	100.0	4.7	94.8	100.0	100.0	9.3
AR(3)	99.8	54.6	86.0	3.9	99.9	55.8	92.7
AR(4)	28.5	100.0	100.0	99.7	4.1	100.0	74.6
AR(5)	100.0	47.2	100.0	50.1	99.9	4.1	100.0
AR(6)	45.2	100.0	19.3	97.5	68.3	100.0	4.1

Table 3: Estimated power of comparison tests for AR models (in percent). For every model 5000 samples with 128 or 256 observations, respectively, have been calculated, and 1000 resamples have been used for each sample.

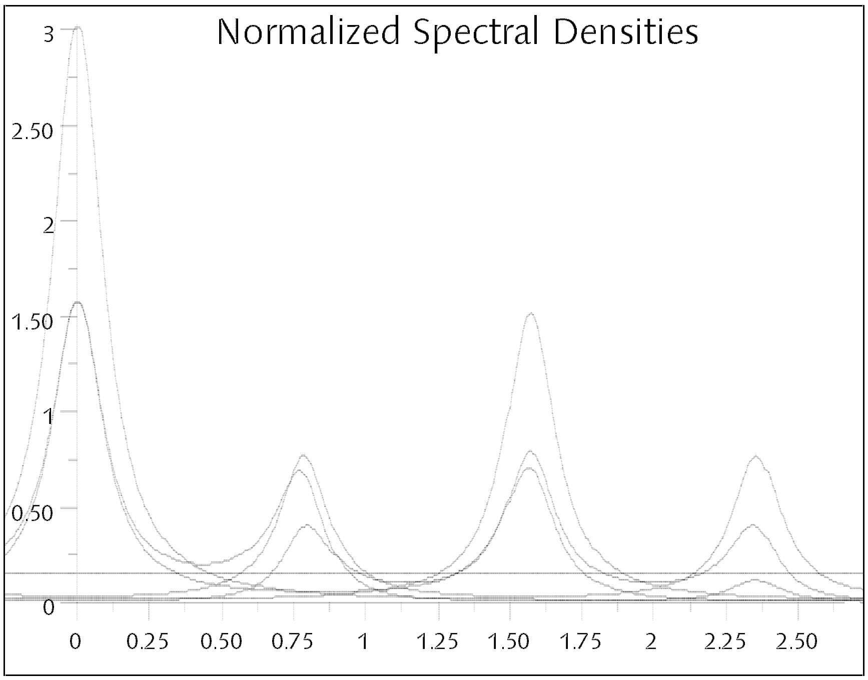

Figure 1: Spectral densities $f^{(i)}(\cdot)$ of the AR(i) processes, $i = 0, \ldots, 6$, with parameters as given in Table 1, divided by $\int_0^{2\pi} f^{(i)}(\lambda) d\lambda$.

model class	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)	AR(7)	AR(8)
p-value (in %)	0.0	0.0	10.8	1.6	2.5	5.5	12.2	37.1	88.4
reject at level									
5%	×	×		×	×				
10%	×	×		×	×	×			

Table 4: Bootstrap goodness-of-fit test for Beveridge Wheat Price Index (1500-1869) using 5000 bootstrap resamples.

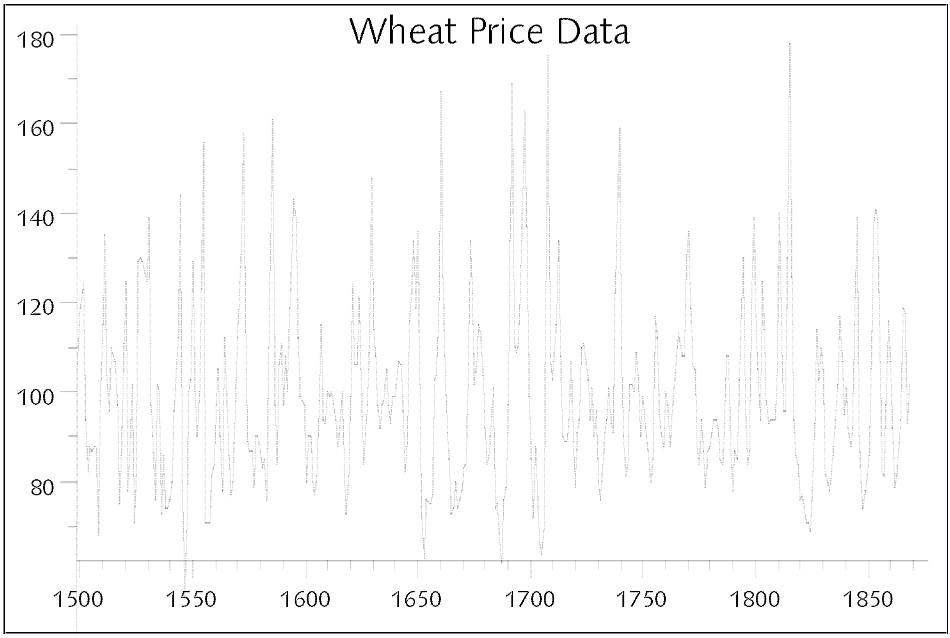

Figure 2: Trend-free Beveridge Wheat Price Index (1500-1869).

subset for the years 1770-1869, assuming these are sufficiently independent so that the test of Theorem 3 may be applied. As model classes for the goodness-of-fit tests we use the AR(i) model classes, $i = 0, \ldots, 8$.

The goodness-of-fit test using 5000 bootstrap resamples produced the results given in Table 4, where the p-values of the tests, i.e. the smallest levels of significance for which the test rejects, are given. We can see that the hypotheses that the data follow a model from a model class with orders 0, 1, 3, 4, and perhaps 5, may be rejected. For a different analysis see Anderson (1971), Sct. 5.9.

Now we compare the data from the years 1500-1599 with those from the years 1770-1869

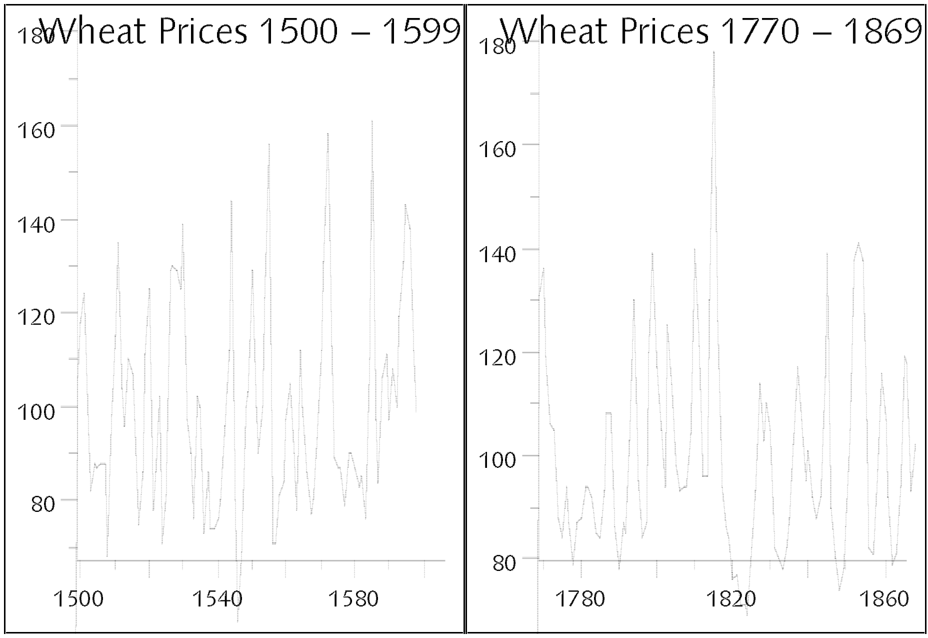

Figure 3: Trend-free Beveridge Wheat Price Index for the years 1500-1599 and 1770-1869.

in order to see whether there can be found a significant difference between their distributions. The data and nonparametric spectral density estimates can be found in Figures 3 and 4, respectively. A test of the hypothesis $H_0^F : f^{(1)}(\lambda)/f^{(2)}(\lambda) = f^{(1)}(\mu)/f^{(2)}(\mu)$ for all $\lambda, \mu \in \Pi$, has been performed as described in Section 5.1, using 5000 bootstrap resamples. As one can conjecture from Figure 4, the comparison test of the hypothesis H_0^F cannot reject even for a level of 10% (in fact, the test statistic corresponds to the 44%-quantile of the distribution of the bootstrap statistic), so no change in the distribution can be found with this method.

6.3 Small sample corrections

The bootstrap methods used in this paper estimate the distributions of the test statistics asymptotically correctly, but for finite samples there can be substantial deviations from these distributions.

One of the reasons for this is that $\hat{f}$ will be a bad estimate of the spectral density f if we have not enough data. The bias of a nonparametric estimate $\hat{f}(\lambda)$ will be large if we have to calculate a kernel estimate by smoothing over periodogram values too far away from λ which do not have an expectation close enough to $f(\lambda)$. To reduce this problem one should use a local bandwidth, or alternatively a prefilter to get approximately equally distributed

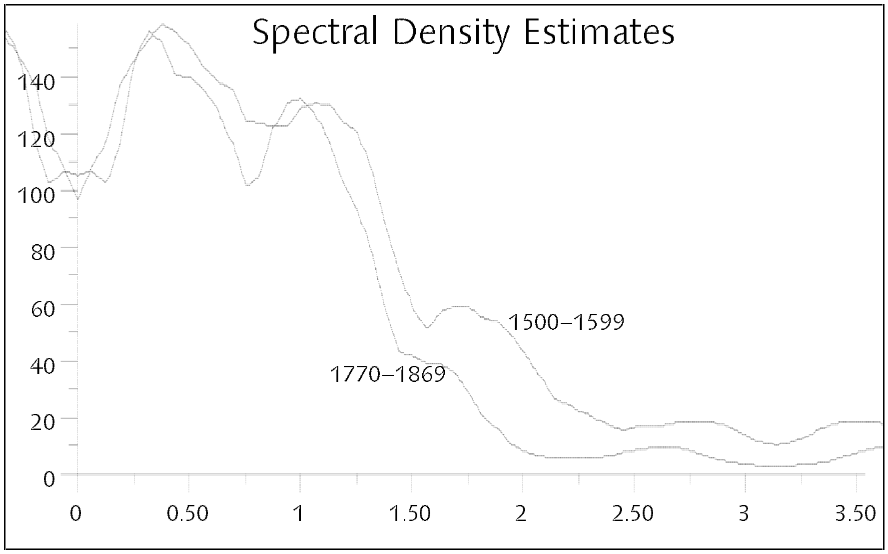

Figure 4: Estimated spectral densities of Beveridge Wheat Price Index for the years 1500-1599 and 1770-1869.

periodogram values before calculating the kernel estimate; this last approach has been used in the above comparison tests with AR-models as prefilters whose orders have been selected by the BIC criterium. The bias of a parametric spectral density estimate, which is used in the goodness-of-fit tests, can be reduced by a bias-correction of the corresponding parameter estimate.

The use of a data taper results in an increase of the variance of the integrated periodograms (cf. Dahlhaus, 1983, Theorem 2) by a factor of $\gamma = TH_{4,T}/H_{2,T}^2$, which should be imitated by a corresponding increase of the variance of the bootstrap periodogram resamples, e.g. by using $I_j^* = \hat{f}_j(\sqrt{\gamma}(\epsilon_j^* - 1) + 1)$ instead of $I_j^* = \hat{f}_j\epsilon_j^*$.

To avoid other difficulties we have used iid exponentially distributed random variables with mean 1 in the simulations of the goodness-of-fit tests (with a data taper and the γ-correction as above), instead of resampling from the estimated residuals $\{\tilde{\epsilon}_j\}$.

Another possibility of correcting the finite sample deviations mentioned above is the use of a parametric bootstrap in the time domain to get a calibration curve for the critical levels of the tests; this method has not been used in the calculations in Sections 6.1 and 6.2, though, because of its huge computational effort. The idea (cp. Loh, 1987) is to imitate the data generation process together with the frequency domain bootstrap in order to find some corrections for systematic distortions of the frequency domain bootstrap. As the true data generation process is unknown, one has to fit a parametric model to the data. If this model decribes the data reasonably well, the estimated corrections should be useful for the true model, too. In contrast to our original data we know the true distribution of the generated (pseudo-) data, and we can calculate the distribution of the test statistic by a Monte Carlo simulation, so we can compare it with the estimated distribution of the test statistic given by the periodogram bootstrap, and we can use these distributions to calculate the calibration curve.

There are several possible ways for calculating a calibration curve, e.g. one can proceed as follows (we restrict ourselves to a single time series of length T, but for the comparison tests a similar method can be applied):

1. Choose a model family Θ which seems to be appropriate to describe the data (for the use in a goodness-of-fit test one should use the model family of the hypothesis). The models should have a form which makes it easy to generate data from it, e.g. an autoregressive model $X_t = \sum_{k=1}^{p} \theta_k X_{t-k} + \eta_t$ for some suitable p.

2. Estimate a model parameter $\hat{\theta} \in \Theta$ from the data.

3. Calculate the corresponding residuals $\{\hat{\eta}_j\}$, e.g. $\hat{\eta}_t := X_t - \sum_{k=1}^{p} \hat{\theta}_k X_{t-k}$, $t = p + 1, \ldots, T$, and center them: $\tilde{\eta}_j := \hat{\eta}_j - \frac{1}{T-p} \sum_{k=p+1}^{T} \hat{\eta}_k$.

4. Repeat steps 4 to 6 S times ($s = 1, \ldots, S$): Generate T pseudo-data from the estimated model $\hat{\theta}$ by resampling $\{\eta_t^{*(s)}\}$ from the estimated innovations $\{\tilde{\eta}_j\}$, e.g. $X_t^{*(s)} = \sum_{k=1}^{p} \hat{\theta}_k X_{t-k}^{*(s)} + \eta_t^{*(s)}$.

5. Treat the pseudo data $\{X_t^{*(s)}\}$ as the real data $\{X_t\}$ are treated in the test procedure, i.e. calculate the test statistic and estimate its distribution with the periodogram bootstrap.

6. Calculate the quantiles $q_\alpha^{(B)}(s)$ of the test statistic for some suitable set of $\alpha \in \mathcal{A} \subset (0, 1)$, e.g. for $\alpha = k/100$, $k = 1, \ldots, 99$.

7. Calculate the mean quantiles $q_\alpha^{(B)} = \frac{1}{S} \sum_{s=1}^{S} q_\alpha^{(B)}(s)$ of the bootstrap distributions and the quantiles q_α of the true distribution of the test statistic (under the estimated model

$\hat{\theta}$), which is estimated by the empirical distribution of the S test statistics calculated in step 5.

8. Calculate the quantile curves by interpolating the point sets $\{(\alpha, q_\alpha^{(B)})\}$ and $\{(\alpha, q_\alpha)\}$, $\alpha \in \mathcal{A}$, respectively.

9. Calculate the calibration curve $(\alpha, \beta(\alpha))$, where $\beta = \beta(\alpha)$ is defined by $q_\beta^{(B)} = q_\alpha$.

If we now want to test a hypothesis on our data set at a level of α with the periodogram bootstrap, we use the $\beta(\alpha)$-quantile of the bootstrap distribution of the test statistic instead of the α-quantile as critical value. The calibration curve maps every requested critical value α to the critical value β that we must use in the bootstrap test to get an improved result.

Other possible calibration methods include the estimation of a bias correction for a given quantile or the estimation of calibration factors for the variances (or even calibration curves for the distributions) of every single periodogram value. However, all these calibration methods require a great computational effort.

Acknowledgement: The simulation programs are based on the extensible data analysis system Voyager developed by G. Sawitzki, M. Diller, F. Friedrich et al. at StatLab Heidelberg. We thank R.v.Sachs for his helpful comments.

Appendix

A. Proof of Lemma 1:

From (2.2) we get

$$\begin{aligned}
\sqrt{T}\Big(\frac{F_{\hat{\theta}}(\lambda)}{F_{\hat{\theta}}(\pi)} - \frac{F_{\theta_0}(\lambda)}{F_{\theta_0}(\pi)}\Big) &= \sqrt{T}(\hat{\theta}-\theta_0)^\top \nabla_\theta\Big(\frac{F_{\theta_0}(\lambda)}{F_{\theta_0}(\pi)}\Big) + \frac{1}{2}\sqrt{T}(\hat{\theta}-\theta_0)^\top \nabla_\theta^2\Big(\frac{F_{\overline{\theta}}(\lambda)}{F_{\overline{\theta}}(\pi)}\Big)(\hat{\theta}-\theta_0) = \\
&= -\sqrt{T}\Big(\frac{1}{2\pi}\int_\Pi((I_T(\alpha) - f_{\theta_0}(\alpha))\nabla_\theta f_{\theta_0}^{-1}(\alpha)d\alpha\Big)^\top \Gamma_0^{-1}\nabla_\theta\Big(\frac{F_{\theta_0}(\lambda)}{F_{\theta_0}(\pi)}\Big) + \\
&\quad + O_P(T^{-1/2}),
\end{aligned}$$

because

$$\nabla_\theta \hat{\Delta}(\theta_0) = \frac{1}{2\pi}\int_\Pi (I_T(\lambda) - f_{\theta_0}(\lambda))\nabla_\theta f_{\theta_0}^{-1}(\lambda)d\lambda$$

follows from Assumptions 1 and 2. The error term is uniform in λ because of Assumption 3. The parameter $\overline{\theta}$ lies between $\hat{\theta}$ and θ_0.

Therefore we get

$$
\begin{aligned}
V(\lambda) = \sqrt{T}(\frac{F_T(\lambda)}{F_T(\pi)} - \frac{F_{\hat\theta}(\lambda)}{F_{\hat\theta}(\pi)}) = & \\
= \; & \sqrt{T}(\frac{1}{F_T(\pi)}\int_\Pi \chi_{[0,\lambda]}(\alpha) I_T(\alpha)d\alpha) - \sqrt{T}\frac{1}{F_{\theta_0}(\pi)}\int_\Pi \chi_{[0,\lambda]}(\alpha) f_{\theta_0}(\alpha)d\alpha + \\
& + \sqrt{T}(\frac{1}{2\pi}\int_\Pi ((I_T(\alpha) - f_{\theta_0}(\alpha))\nabla_\theta f_{\theta_0}^{-1}(\alpha)d\alpha)^\top \Gamma_0^{-1}\nabla_\theta(\frac{F_{\theta_0}(\lambda)}{F_{\theta_0}(\pi)}) + O_P(T^{-1/2}) = \\
= \; & \sqrt{T}\int_\Pi \phi_\lambda(\alpha) I_T(\alpha)d\alpha + \sqrt{T}\int_\Pi \chi_{[0,\lambda]}(\alpha)(\frac{f(\alpha)}{\int_\Pi f(\gamma)d\gamma} - \frac{f_{\theta_0}(\alpha)}{\int_\Pi f_{\theta_0}(\gamma)d\gamma})d\alpha + \\
& + \sqrt{T}(\frac{1}{2\pi}\int_\Pi ((I_T(\alpha) - f_{\theta_0}(\alpha))\nabla_\theta f_{\theta_0}^{-1}(\alpha)d\alpha)^\top \Gamma_0^{-1}\nabla_\theta(\frac{F_{\theta_0}(\lambda)}{F_{\theta_0}(\pi)}) + O_P(T^{-1/2}) = \\
= \; & \overline{V}(\lambda) + O_P(T^{-1/2}),
\end{aligned}
$$

because $\int_\Pi I_T(\alpha)d\alpha = \int_\Pi f(\alpha)d\alpha + O_P(T^{-1/2})$. The remainder terms are uniform in λ. The representation of $\overline{V}(\lambda)$ under H_0 follows with the arguments given prior to Lemma 1.

B. Proof of Lemmas 2 and 4:

From Lemma 6 and Lemma 7 of Dahlhaus (1983) and by the same arguments as in Dahlhaus and Janas (1996), Lemma 1, we get for functions $\psi^{(i)}$ bounded and of bounded total variation:

$$
\begin{aligned}
T^{1/2}E\int_\Pi \psi^{(1)}(\alpha)I_T(\alpha)d\alpha &= T^{1/2}\int_\Pi \psi^{(1)}(\alpha)f(\alpha)d\alpha + o(1), \\
TCum(\int_\Pi \psi^{(1)}(\alpha)I_T(\alpha)d\alpha, \int_\Pi \psi^{(2)}(\alpha)I_T(\alpha)d\alpha) &= 2\pi\int_\Pi \psi^{(1)}(\alpha)\psi^{(2)}(\alpha)f^2(\alpha)d\alpha + \\
+\frac{\kappa_4}{\sigma^4}(\int_\Pi \psi^{(1)}(\alpha)f(\alpha)d\alpha)(\int_\Pi \psi^{(2)}(\alpha)f(\alpha)d\alpha) &+ o(1), \\
T^{r/2}Cum(\int_\Pi \psi^{(1)}(\alpha)I_T(\alpha)d\alpha, \ldots, \int_\Pi \psi^{(r)}(\alpha)I_T(\alpha)d\alpha) &= o(1) \quad \text{for } r \geq 3.
\end{aligned}
$$

The remainder terms are uniform for all functions $\psi^{(i)}$ which are uniformly bounded and of uniformly bounded variation; this is the case for the functions $\{\phi_\lambda + \omega_\lambda, \lambda \in \Pi\}$ by Assumption 3.

Therefore the moments of $\overline{V}(\lambda)$ are

$$
\begin{aligned}
E\overline{V}(\lambda) &= ET^{1/2}\int_\Pi (\phi_\lambda(\alpha) + \omega_\lambda(\alpha))I_T(\alpha)d\alpha + \sqrt{T}\int_\Pi \chi_{[0,\lambda]}(\alpha)(\frac{f(\alpha)}{\int_\Pi f(\gamma)d\gamma} - \frac{f_{\theta_0}(\alpha)}{\int_\Pi f_{\theta_0}(\gamma)d\gamma})d\alpha = \\
&= \sqrt{T}\int_\Pi \chi_{[0,\lambda]}(\alpha)(\frac{f(\alpha)}{\int_\Pi f(\gamma)d\gamma} - \frac{f_{\theta_0}(\alpha)}{\int_\Pi f_{\theta_0}(\gamma)d\gamma})d\alpha + o(1),
\end{aligned}
$$

since $\int_\Pi (\phi_\lambda(\alpha) + \omega_\lambda(\alpha))f(\alpha)d\alpha = 0$, and for the same reason

$$
Cov(\overline{V}(\lambda), \overline{V}(\mu)) = Cov(T^{1/2}\int_\Pi (\phi_\lambda(\alpha) + \omega_\lambda(\alpha))I_T(\alpha)d\alpha, T^{1/2}\int_\Pi (\phi_\mu(\alpha) + \omega_\mu(\alpha))I_T(\alpha)d\alpha) =
$$

$$
\begin{aligned}
&= 2\pi \int_\Pi f^2(\alpha)(\phi_\lambda(\alpha)+\omega_\lambda(\alpha))(\phi_\mu(\alpha)+\omega_\mu(\alpha))d\alpha + \\
&\quad + \frac{\kappa_4}{\sigma^4}(\int_\Pi f(\alpha)(\phi_\lambda(\alpha)+\omega_\lambda(\alpha))d\alpha)(\int_\Pi f(\alpha)(\phi_\mu(\alpha)+\omega_\mu(\alpha))d\alpha) + o(1) = \\
&= 2\pi \int_\Pi f^2(\alpha)(\phi_\lambda(\alpha)+\omega_\lambda(\alpha))(\phi_\mu(\alpha)+\omega_\mu(\alpha))d\alpha + o(1).
\end{aligned}
$$

The proof for the higher order cumulants is obvious. Lemma 4 is proved similarly using $\psi_\lambda(\cdot)$ as integrand; here the κ_4-term does not vanish, though. The convergence to a normal distribution follows from the convergence of the cumulants to the cumulants of the normal distribution.

C. Proof of Lemmas 3 and 5:

As in (2.2) we get

$$
T^{1/2}(\hat{\theta}^* - \theta^*) = -\Gamma_0^{*-1} T^{1/2} \nabla_\theta \hat{\Delta}^*(\theta^*) + O_{P^*}(T^{-1/2}),
$$

where $\Gamma_0^* = \nabla_\theta^2 \Delta^*(\theta^*) \to \Gamma_0$ a.s.

By the same arguments as in Appendix A we can see that $V^*(\lambda) = \overline{V}^*(\lambda) + O_{P^*}(T^{-1/2})$ with remainder term uniform in λ, where

$$
\begin{aligned}
\overline{V}^*(\lambda) &= T^{1/2}\frac{2\pi}{T}\sum_{j=1}^{n}(\hat{\phi}_\lambda(\lambda_j)+\hat{\omega}_\lambda(\lambda_j))I_j^* + \\
&\quad + T^{1/2}\frac{2\pi}{T}\sum_{j=1}^{n}\chi_{[0,\lambda]}(\lambda_j)\Big(\frac{\hat{f}(\lambda_j)}{\frac{2\pi}{T}\sum_{k=1}^{n}\hat{f}(\lambda_k)} - \frac{f_{\theta^*}(\lambda_j)}{\frac{2\pi}{T}\sum_{k=1}^{n} f_{\theta^*}(\lambda_k)}\Big).
\end{aligned}
$$

Here $\hat{\phi}_\lambda$ and $\hat{\omega}_\lambda$ are defined as ϕ_λ and ω_λ, respectively, but with Γ_0 replaced by Γ_0^*, θ_0 by θ^*, f by $\hat{f}$, and integrals by sums.

The proof of Lemma 3 now follows as in Dahlhaus and Janas (1996), cp. their Lemma 4, and Appendix B, using the uniform convergences $\hat{\phi}_\lambda(\alpha) \to \phi_\lambda(\alpha)$ and $\hat{\omega}_\lambda(\alpha) \to \omega_\lambda(\alpha)$ for all λ and α (Assumptions 3 and 6).

Similarly Lemma 5 follows from an expansion as in (4.3), using the theorem in Janas and Dahlhaus (1994). Note that the additional term c^* in the definition of I_j^* has no influence on the conditional mean of the bootstrap periodogram.

D. Proof of Theorems 1 and 2:

We have seen (Lemma 3) that under H_0 the finite-dimensional distribution of $(\overline{V}(\mu_1), \ldots, \overline{V}(\mu_k))$ is imitated by the conditional distribution of $(\overline{V}^*(\mu_1), \ldots, \overline{V}^*(\mu_k))$ for all fixed $\mu_1, \ldots, \mu_k \in \Pi$, and from Lemma 1 and Appendix C it follows that this holds for the statistics $V(\mu_i)$ and $V^*(\mu_i)$, respectively, too. Theorem 1 will follow if the distribution of the supremum of the absolute values over all λ is estimated consistently by the conditional distribution of the corresponding bootstrap statistic; this will be a consequence of the continuous mapping theorem and the fact that the processes $\{\overline{V}(\lambda)\}$ and $\{\overline{V}^*(\lambda\}$, $\lambda \in \Pi$, converge (almost surely) to the same Gaussian limit process (in the sense of Dahlhaus, 1988).

The convergence of the empirical spectral process $\{\overline{V}(\lambda)\}$ follows under our assumptions from Dahlhaus (1988), and for the bootstrap process $\{\overline{V}^*(\lambda)\}$ it follows from the convergence of the finite-dimensional distributions to the Gaussian distribution with the same covariances as for the original statistics, and from the equicontinuity of the process. The equicontinuity can be shown similarly as in Pollard (1984), Lemma VII.4(15), where it is proved for the empirical process of independent and identically distributed random variables. The necessary conditions on the covering numbers follow from Assumption 3. As the bootstrap periodogram values I_j^* are independent, though not identically distributed, the result carries over to our case with some modifications: The mean of the random variables, $P\phi$ in Pollard (1984), is replaced by $\frac{1}{n}\sum_{j=1}^{n}(\hat{\phi}_\lambda(\lambda_j) + \hat{\omega}_\lambda(\lambda_j))\hat{f}_j$, and the mean of the empirical distribution, $P_n\phi$, is replaced by $\frac{1}{n}\sum_{j=1}^{n}(\hat{\phi}_\lambda(\lambda_j) + \hat{\omega}_\lambda(\lambda_j))I_j^*$ (see (3.1)). As $\frac{1}{n}\sum_{j=1}^{n} I_j^{*2}$ is of order $O(1)$ a.s., the equicontinuity follows (see also Dahlhaus, 1988).

The proof of Theorem 2 follows essentially the same way as the proof of Theorem 1, replacing $\hat{\phi}_\lambda + \hat{\omega}_\lambda$ by $\hat{\psi}_\lambda$, because the factor $T^{1/2}c^*$ in the additional term $T^{1/2}c^*\frac{2\pi}{T}\sum_{j=1}^{n}\hat{\psi}_\lambda(\lambda_j)\hat{f}_j$, which is not covered by the above considerations, is of order $O_{P^*}(1)$, and $\frac{2\pi}{T}\sum_{j=1}^{n}\hat{\psi}_\lambda(\lambda_j)\hat{f}_j \to \int_\Pi \psi_\lambda(\alpha)f(\alpha)d\alpha =: z(\lambda)$ uniformly for all λ. As $z(\lambda)$ is (uniformly) continuous by Assumptions 3 and 4, this leads to the equicontinuity of the summed bootstrap periodogram generated by the wild bootstrap.

Similar results for Gaussian time series have been obtained by Nordgaard (1995) who shows the convergence of the bootstrap empirical process for a related bootstrap method.

E. Proof of Theorems 3 and 4:

We write for short

$$D_T^{(i)}(\lambda) := (\int_\Pi \chi_{[0,\lambda]}(\alpha) I^{(i)}(\alpha)d\alpha)/(\int_\Pi I^{(i)}(\alpha)d\alpha) - (\int_\Pi \chi_{[0,\lambda}(\alpha) f^{(i)}(\alpha)d\alpha)/(\int_\Pi f^{(i)}(\alpha)d\alpha),$$

and similarly

$$D_T^{(i)*}(\lambda) := (\frac{1}{n_i}\sum_{j=1}^{n_i} \chi_{[0,\lambda]}(\lambda_j) I_j^{(i)*})/(\frac{1}{n_i}\sum_{j=1}^{n_i} I_j^{(i)*}) - (\frac{1}{n_i}\sum_{j=1}^{n_i} \chi_{[0,\lambda]}(\lambda_j)\hat{g}^{(i)}(\lambda_j))/(\frac{1}{n_i}\sum_{j=1}^{n_i} \hat{g}^{(i)}(\lambda_j)).$$

Theorem 3 can be shown similarly as Theorems 1 and 2 by first approximating $D_T^{(i)}(\lambda)$ and $D_T^{(i)*}(\lambda)$ by some suitable statistics ${D'}_T^{(i)}(\lambda)$ and ${D'}_T^{(i)*}(\lambda)$, proving the convergence of the multivariate distribution of $\{T_i^{1/2}{D'}_T^{(i)}(\lambda_k), k \in \mathcal{M}\}$ and $\{T_i^{1/2}{D'}_T^{(i)*}(\lambda_k), k \in \mathcal{M}\}$ (a.s.) to the same limit distribution (for arbitrary finite sets $\mathcal{M} \subset \mathbf{N}$), and finally showing the convergence of $\{T_i^{1/2}{D'}_T^{(i)}(\lambda)\}$ and $\{T_i^{1/2}{D'}_T^{(i)*}(\lambda)\}$ to the same Gaussian limit process by proving equicontinuity. By the usual argument for a comparison test for independent Gaussian random variables the convergence results for $\sqrt{T_1T_2/(T_1+T_2)}({D'}_T^{(1)}(\lambda) - {D'}_T^{(2)}(\lambda))$, their bootstrap counterpart, and the suprema follow.

The first two points can be handled as in Dahlhaus and Janas (1996), because $D_T^{(i)}(\lambda)$ is a ratio statistic; the additional assumptions in Dahlhaus and Janas (1996) are unnecessary in our case as we do not need a higher order approximation of the statistic's distribution by the bootstrap. As the variance of $T_i^{1/2}{D'}_T^{(i)}(\lambda)$ asymptotically only depends on the normalized spectral densities (see Dahlhaus and Janas, 1996, Theorem 4), we have the same variances under H_0^F for $i=1$ and $i=2$, and $\sqrt{T_1T_2/(T_1+T_2)}({D'}_T^{(1)}(\lambda) - {D'}_T^{(2)}(\lambda))$ is asymptotically normally distributed; the equicontinuity follows from the equicontinuity of the statistics of the two independent processes (cp. Anderson, 1993, Sct. 5). As the same arguments hold for the bootstrap processes $\{D_T^{(i)*}(\lambda)\}$ (cp. Appendix D), we get Theorem 3 similarly as Theorems 1 and 2.

The proof of Theorem 4 is exactly the same except that now no approximations like ${D'}_T^{(i)}(\lambda)$ and ${D'}_T^{(i)*}(\lambda)$ are needed, and the asymptotic distributions of the processes depend on the fourth cumulants $\kappa_4^{(i)}$, which is the reason for the appearance of $c = c(\kappa_4^{(1)}, \kappa_4^{(2)})$ in the asymptotic distribution ($c = 1$ if $\kappa_4^{(1)} = \kappa_4^{(2)}$); we omitt further details.

References

Akaike,H.(1972). Use of an information theoretic quantity for statistical model identification. *Proc.5th Hawaii Inter. Conference on System Sciences* 249-250.

Anderson,T.W.(1971). The Statistical Analysis of Time Series. Wiley, New York.

Anderson,T.W.(1993). Goodness of fit tests for spectral distributions. *Ann.Statist.* **21** 830-847.

Anderson,T.W.(1997). Goodness-of-fit tests for autoregressive processes. *J.Time Ser.Anal.* **18** 321-339.

Bartlett, M.S.(1954). Problèmes de l'Analyse Spectrale des Séries Temporelles Stationnaires. *Publ.Inst.Statist.Univ.Paris* **III-3** 119-134.

Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions. *Ann.Stat.* **16** 1709-1722.

Brillinger, D.R.(1981). Time Series. Data Analysis and Theory. McGraw-Hill, N.Y.

Chen, H., Romano, J.P. (1997). Bootstrap Goodness of Fit Tests in the Frequency Domain. Preprint.

Chen, Z.-G., Hannan, E.J. (1980). The Distribution of Periodogram Ordinates. *J.Time Ser.Anal.* **1** 73-82.

Dahlhaus,R.(1983). Spectral analysis with tapered data. *J. Time Ser. Anal.* **4** 163-175.

Dahlhaus,R.(1985). On the Asymptotic Distribution of Bartlett's U_p-Statistic. *J. Time Ser. Anal.* **6** 213-227.

Dahlhaus,R.(1988). Empirical Spectral Processes And Their Application To Time Series Analysis, *Stoch. Proc. Appl.*. **30** 69-83.

Dahlhaus,R., Janas, D. (1996): *A frequency domain bootstrap for ratio statistics in time series analysis*, Ann.Stat., vol. 24, p. 1934-1963.

Diggle, P.J., Fisher, N.I.(1991). Nonparametric comparison of cumulative periodograms. *J.Roy.Statist.Soc.Ser.C (Appl. Statist.)* **40** 423-434.

Franke, J., Härdle, W.(1992). On bootstrapping kernel spectral estimates. *Ann.Statist.* **20** 121-145.

Grenander, U., Rosenblatt, M.(1956). Statistical Analysis of Stationary Time Series. Almquist and Wiksell, Stockholm.

Janas, D., Dahlhaus, R. (1994). A frequency domain bootstrap for time series. In: Computationally Intensive Statistical Methods. Proc. of the 26th Symp. on the Interface (J.Sall, A.Lehman, eds.). 423-425. Interface Foundation of North America, Fairfax Station, VA.

Janas,D., von Sachs,R.(1995). Consistency for non-linear functions of the periodogram of tapered data. *J.Time Ser.Anal.* **16** 585-606.

Kreiss, J.-P., Franke, J.(1992). Bootstrapping Stationary Autoregressive Moving-Average Models. *J.Time Ser.Anal.* **13** 297-317.

Loh,W.-Y.(1987). Calibrating Confidence Coefficients. *J.Amer. Statist. Assoc.* **82** 155-162.

Nordgaard,A.(1995). Resampling the spectral measure I: The parametric case. preprint LiU-MAT-R-1995-01, Department of Mathematics, Linköping University.

Paparoditis, E.(1995). Spectral Density Based Goodness - Of - Fit Tests for Time Series Models. preprint.

Parzen, E.(1983). Autoregressive spectral estimation. In: Handbook of Statistics 3, D.R. Brillinger and P.R.Krishnaiah (eds.), North Holland, Amsterdam.

Pollard,D.(1984). Convergence of Stochastic Processes. Springer, N.Y.