The Blockwise Bootstrap for General Parameters of a Stationary Time Series

by

Peter Bühlmann and
Hans R. Künsch

Research Report No. 70
May 1993

Seminar für Statistik

Eidgenössische Technische Hochschule (ETH)

CH-8092 Zürich

Switzerland

The Blockwise Bootstrap for General Parameters of a Stationary Time Series

Peter Bühlmann and
Hans R. Künsch

Seminar für Statistik
ETH Zentrum
CH-8092 Zürich, Switzerland

21 May 1993

Abstract

We study the blockwise bootstrap of Künsch (1989) for a statistic which estimates a parameter of the entire distribution of a stationary time series. Because such a statistic is not symmetric in the observations, one should not simply resample blocks of the original data. When the parameter is the spectral distribution function or an ARMA parameter, the statistic is a symmetric function of all shifts of the sample extended suitably. Then we can resample blocks of shift indices, and the theory is basically the same as for a symmetric statistic. In other cases the statistic is a symmetric function of m-tuples of consecutive data where m increases with sample size. Then one can resample blocks of these m-tuples. But the increasing m makes the theory more delicate. We show validity of the bootstrap in two generic examples of spectral estimators, thereby extending results of Politis and Romano (1992).

Keywords: Resampling; Nonparametric spectral estimation; Autoregressive spectral estimation; ARMA-processes; Estimators defined by functionals; Central limit theorem; Strong mixing sequence; Martingale differences.

Running headline: The blockwise bootstrap

1 Introduction

Independence of the observations is crucial for Efron's (1979) bootstrap. For time series data a modification is needed. One such modification which does not assume a special structure like ARMA– or Markov processes is the blockwise bootstrap of Künsch (1989). It proceeds by resampling blocks of ℓ consecutive observations from the original sample. Different blocks are conditionally independent, and this causes problems when the statistic T_n to be bootstrapped is not symmetric in the observations. Namely between independent blocks there may be large jumps. So if T_n looks for instance at all pairs of observations with a fixed lag distance and we compute T_n for the bootstrap sample, it also looks at many unrelated pairs and is thus biased. In order to avoid this Künsch (1989) proposed the following: Assume that T_n depends on some m-dimensional marginal with m fixed, i.e. T_n is a symmetric function of $X_1', \ldots, X_{n-m+1}'$ where $X_t' = (X_t, X_{t+1}, \ldots, X_{t+m-1})$. Then one should apply the blockwise resampling not to the original data X_t, but rather to X_t'. Politis and Romano (1992) took this idea up, called it the "block of blocks" bootstrap and investigated what happens when m is not fixed, but increases slowly with m. They mainly considered the nonparametric spectral estimator which averages short periodograms (Welch, 1967). They showed that there the block of blocks bootstrap is still consistent. But their result has some unpleasant and unnatural features. First they require that T_n is not a symmetric function of all X_t''s but only of those separated by multiples of a lag L where L is of the same order as m. This modification of T_n is in general not asymptotically negligible and its effect can in some cases be rather large. Moreover they require that we resample only blocks which are separated by a distance of the same order as the block length ℓ.

The aim of this paper is to develop a general theory for bootstrapping statistics which estimate a parameter of the distribution of the entire underlying process. We distinguish two types of such statistics. The first one extends the sample by e.g. the arithmetic mean and then uses the average of shifts of this extended sample. Examples are estimators of ARMA (p, q) - parameters for $q > 0$ and of the spectral distribution function. For these statistics we can bootstrap simply by resampling blocks of shifts of the extended sample. The second type is the one described above where T_n is a functional of a marginal whose dimension m increases with n. For such statistics we show that the block of blocks bootstrap typically works without the restrictions imposed by Politis and Romano (1992). The essential condition we need is that the block length ℓ in the resampling must increase faster than m. This is a natural condition since X_t' and X_s' are strongly dependent for $|t - s| \leq m$ and ℓ must be big enough to capture all dependence. We illustrate this by studying two spectral estimators in detail, the nonparametric estimator and the AR-spectrum.

2 Estimation of parameters of the entire distribution

Let $(X_t)_{t \in \mathbb{Z}}$ be stationary and for simplicity real valued. Define

$$F^m = \mathcal{L}(X_1, \ldots, X_m) \quad \text{(m-dimensional marginal)}$$

$$F^{\infty} \quad = \quad \mathcal{L}(X_t, t \in \mathbb{Z}) \qquad \text{(whole distribution)}.$$

Assume that we are interested in a real-valued parameter θ depending on the whole distribution, i.e. $\theta = T(F^{\infty})$. Examples are

The spectral density at a fixed frequency λ:

$$\theta = f(\lambda) = \mathrm{Var}(X_1) + 2\sum_{k=1}^{\infty} \mathrm{Cov}(X_1, X_{1+k}) \cos(k\lambda) \tag{2.1}$$

The spectral distribution function at a fixed frequency:

$$\theta = F(\lambda) = \int_0^{\lambda} f(\omega) d\omega = \lambda \mathrm{Var}(X_1) + 2\sum_{k=1}^{\infty} \mathrm{Cov}(X_1, X_{1+k}) \sin(k\lambda)/k \tag{2.2}$$

The parameter of an approximating MA(1)- processes:

$$\theta = \mathrm{argmin}_{\alpha \in (-1,1)} E[(X_1 - E[X_1] - \sum_{j=1}^{\infty} \alpha^j (X_{1-j} - E[X_1]))^2]. \tag{2.3}$$

In general there are several ways to estimate θ from observations $X_1, \ldots, X_n$ depending on the behavior of the functional T. In some cases direct estimation is possible: we extend the observations by e.g. the arithmetic mean

$$X_t^{\mathrm{ext}} = \begin{cases} X_t & \text{if} \quad 1 \le t \le n \\ \overline{X} = n^{-1}\sum_{s=1}^n X_s & \text{if} \quad t < 1 \text{ or } t > n \end{cases}$$

and compute

$$\theta_n = T(n^{-1} \sum_{t=0}^{n-1} \Delta(\tau_t X^{\mathrm{ext}})). \tag{2.4}$$

Here $\Delta(x)$ is the point mass at $x \in \mathbb{R}^{\mathbb{Z}}$ and τ_t is the shift operator $(\tau_t x)_s = x_{s+t}$. By this procedure we obtain in Example (2.1)

$$\theta_n = n^{-1} \sum_{t=1}^{n} (X_t - \overline{X})^2 + 2 \sum_{k=1}^{n-1} n^{-1} \sum_{t=1}^{n-k} (X_t - \overline{X})(X_{t+k} - \overline{X}) \cos(k\lambda) \tag{2.5}$$

the periodogram, in Example (2.2)

$$\theta_n = n^{-1} \sum_{t=1}^{n} (X_t - \overline{X})^2 \lambda + 2 \sum_{k=1}^{n-1} n^{-1} \sum_{t=1}^{n-k} (X_t - \overline{X})(X_{t+k} - \overline{X}) \frac{\sin(k\lambda)}{k} \tag{2.6}$$

the empirical spectral distribution function, and in Example (2.3)

$$\theta_n = \mathrm{argmin}_{\alpha \in (-1,1)} \sum_{t=2}^{n} (X_t - \overline{X} - \sum_{j=1}^{t-1} \alpha^j (X_{t-j} - \overline{X}))^2, \tag{2.7}$$

the least-squares estimator, respectively. It is well known that (2.6) and (2.7) are consistent for a broad class of distributions F^{∞}, but (2.5) is never consistent. In order to estimate the spectral density consistently, we can for instance approximate θ by a sequence of parameters θ^m depending only on F^m:

$$\theta^m = T^m(F^m), \quad \theta^m \underset{m\to\infty}{\longrightarrow} \theta$$

and then plug in the empirical marginal F_n^m instead of F^m:

$$\theta_n^m = T^m((n-m+1)^{-1} \sum_{t=0}^{n-m} \Delta(\tau_t(X_1, \ldots, X_m))). \tag{2.8}$$

Typically θ_n^m is consistent when we choose $m = m(n)$ such that $1/m + m/n \to 0$. In example (2.1) we can choose either

$$\theta^m = \mathrm{Var}(X_1) + 2\sum_{k=1}^{m-1} w(k,m)\mathrm{Cov}(X_1, X_{1+k})\cos(k\lambda) \tag{2.9}$$

or

$$\theta^m = \sigma_{m-1}^2 |1 - \sum_{k=1}^{m-1} a_{k,m-1}\exp(i\lambda k)|^{-2} \tag{2.10}$$

where

$$(a_{1,m-1}, \ldots, a_{m-1,m-1}) = \mathrm{argmin}_{a_i} E[(X_m - E[X_1] - \sum_{j=1}^{m-1} a_j(X_{m-j} - E[X_1]))^2]$$

and σ_{m-1}^2 is the minimum of the function on the right hand side. With (2.9) θ_n^m becomes the usual lag weight estimator whereas with (2.10) θ_n^m is the spectrum of a fitted $AR(m-1)$ – process.

The above example shows that in general there may be many approximations (θ^m) of θ. This is of course quite obvious. Note however that there is a second, less obvious cause of non-uniqueness. Namely F^m is stationary, but F_n^m is not because of boundary effects. Hence there may be two functionals T^m and U^m such that $T^m(F^m) = U^m(F^m)$ but $T^m(F_n^m) \neq U^m(F_n^m)$.

For instance if we take

$$U(F^m) = E|\sum_{k=1}^{m} v(k)(X_k - E[X_k])e^{i\lambda k}|^2 / \sum_{k=1}^{m} v(k)^2 \tag{2.11}$$

and $T^m(F^m)$ as the right-hand side of (2.9), then $U^m(F^m) = T^m(F^m)$ provided $w(k,m) = \sum_{j=1}^{m-k} v(j)v(j+k)/\sum_{j=1}^{m} v(j)^2$. Now $U^m(F_n^m)$ is the tapered and averaged periodogram which is asymptotically equivalent, but not equal to the lag weight estimator, cf. Welch (1967).

In order to obtain asymptotic variances and asymptotic normality of these estimators, the general strategy is to linearize the estimator and then to use a central limit theorem for sums of dependent random variables. In the case of direct estimation (2.4) we expect that

$$\theta_n - \theta = n^{-1} \sum_{t=0}^{n-1} \phi(\tau_t X) + R_n \tag{2.12}$$

with a negligible remainder term R_n. This can be verified in examples (2.6) and (2.7). For (2.6) we have

$$\theta_n = n^{-1}\sum_{t=1}^{n}(X_t - \overline{X})\left((X_t - \overline{X})\lambda + \sum_{\substack{k=-t+1\\k\neq 0}}^{n-t}(X_{t+k} - \overline{X})\sin(k\lambda)/k\right)$$

$$\approx n^{-1}\sum_{t=1}^{n}(X_t - E[X_t])\left((X_t - E[X_t])\lambda + \sum_{\substack{k=-t+1\\k\neq 0}}^{n-t}(X_{t+k} - E[X_t])\sin(k\lambda)/k\right).$$

Using the Cramér-representation $X_t - E[X_t] = \int_{-\pi}^{\pi} e^{i\omega t}Z(d\omega)$ it can be shown that in quadratic mean as $m \to \infty$

$$(X_t - E[X_t])\lambda + \sum_{\substack{k=-m\\k\neq 0}}^{m}(X_{t+k} - E[X_t])\sin(k\lambda)/k \longrightarrow \frac{1}{2}\int_{-\lambda}^{\lambda} e^{it\omega}Z(d\omega)$$

cf. e.g. Doob (1953), p. 482 – 483. Hence in the example (2.6), (2.12) holds with

$$\phi(X) = \frac{1}{2}(X_0 - E[X])\int_{-\lambda}^{\lambda} Z(d\omega) - \theta = \frac{1}{2}\int_{-\pi}^{\pi}\int_{-\lambda}^{\lambda} Z(d\omega_1)Z(d\omega_2) - \theta.$$

It is easily checked that $E[\phi] = 0$.

For example (2.7) the usual linearization argument gives (2.12) with

$$\phi(X) = e_1 u_1/E[u_t^2]$$

where $e_t = \sum_{j=0}^{\infty}\theta^j(X_{t-j} - E[X_{t-j}])$ and $u_t = \sum_{j=1}^{\infty} j\theta^{j-1}(X_{t-j} - E[X_{t-j}])$.

In the other case, where we first approximate θ by θ^m, we obtain by linearization

$$\theta_n^m - \theta^m = (n-m+1)^{-1}\sum_{t=0}^{n-m}\phi_m(\tau_t(X_1,\ldots,X_m)) + R_{m,n} \tag{2.13}$$

with usually negligible remainder $R_{m,n}$. In the case of spectral density estimation, we have for (2.9)

$$\begin{aligned}&\phi_m(X_1,\ldots,X_m) &(2.14)\\ &= (X_m - E[X_t])\left((X_m - E[X_t]) + 2\sum_{k=1}^{m-1} w(k,m)(X_{m-k} - E[X_t])\cos(k\lambda)\right) - \theta^m,\end{aligned}$$

and for (2.10) (see Berk, 1974)

$$\phi_m(X_1,\ldots,X_m) = -\gamma_m(\lambda)^T R(m-1)^{-1}\underline{X}_1^{m-1}(X_m - \sum_{j=1}^{m-1} a_j X_{m-j}). \tag{2.15}$$

Here (a_j) are the coefficients of the innovations e_t in the Wold representation, i.e. $e_t = X_t - \sum_{j=1}^{\infty} a_j X_{t-j}$, and

$$\underline{X}_1^{m-1} = (X_{m-1}, X_{m-2}, \ldots, X_1)^T$$

$$
\begin{aligned}
R(m) &= (\mathrm{Cov}(X_t, X_s);\ 1 \le t, s \le m) \\
\gamma_m(\lambda) &= 2f(\lambda)^2 \sigma_e^{-2} (a(\lambda)(\cos(\lambda), \cos(2\lambda), \ldots, \cos((m-1)\lambda))^T \\
&\quad + b(\lambda)(\sin(\lambda), \ldots, \sin((m-1)\lambda))^T) \\
a(\lambda) &= 1 - \sum a_j \cos(j\lambda), b(\lambda) = -\sum a_j \sin(j\lambda) \\
\sigma_e^2 &= \mathrm{Var}(e_t).
\end{aligned}
$$

The non-uniqueness we discussed before appears also in the linearization (2.13). Namely

$$
(n-m+1)^{-1} \sum_{t=0}^{n-m} \phi_m(X_{t+1}, \ldots, X_{t+m})
$$

and

$$
(n-m+1)^{-1} \sum_{t=0}^{n-m} \overline{\phi}_m(X_{t+1}, \ldots, X_{t+m})
$$

are typically equivalent if

$$
\overline{\phi}_m(X_1, \ldots, X_m) = \phi_m(X_1, \ldots, X_m) + g(X_1, \ldots, X_{m-1}) - g(X_2, \ldots, X_m) \tag{2.16}
$$

for arbitrary g. This phenomenon has been discussed in Künsch (1984) in connection with the definition of the influence function for time series. For instance when we start with (2.11) instead of (2.9) we are naturally lead to

$$
\phi_m(X_1, \ldots, X_m) = (\sum_{k=1}^{m} v(k)^2)^{-1} | \sum_{k=1}^{m} (X_k - E[X_k]) v(k) e^{i\lambda k}|^2 - \theta^m \tag{2.17}
$$

It can be easily verified that between the ϕ_m's of (2.14) and (2.17) we have relation (2.16).

The difference between (2.12) and (2.13) is that typically $\lim_{m\to\infty} \phi_m(X_{2-m}, X_{3-m}, \ldots, X_1)$ does not exist. This is easily verified in the case of (2.14) and (2.15). So in these examples it is really necessary to approximate θ by θ^m.

This difference shows also up in the variances of the linearized parts of $\theta_n - \theta$ and $\theta_n^m - \theta^m$ respectively. In the case of (2.12) we typically have

$$
\mathrm{Var}(\theta_n - \theta - R_n) \sim n^{-1} \sigma^2 \tag{2.18}
$$

with

$$
\sigma^2 = \sum_{t=-\infty}^{\infty} E[\phi(X)\phi(\tau_t X)] \in (0, \infty) \tag{2.19}
$$

whereas in the case of (2.13)

$$
\mathrm{Var}(\theta_n^m - \theta^m - R_{m,n}) \sim m n^{-1} \sigma^2 \tag{2.20}
$$

with

$$
\sigma^2 = \lim\ m^{-1} \sum_{t=-\infty}^{\infty} \mathrm{Cov}\left(\phi_m(X_1, \ldots, X_m), \phi_m(\tau_t(X_1, \ldots, X_m))\right). \tag{2.21}
$$

That the sum of covariances in (2.21) grows with m, can be due to different reasons. It can happen that the variance of the ϕ_m's is stable, but their correlations grow with m because there is more and more overlap:

$$
\lim_{m\to\infty} E[\phi_m(X_1, \ldots, X_m)^2] = c \tag{2.22}
$$

$$\lim_{m\to\infty, t/m\to x} Corr(\phi_m(X_1,\ldots,X_m),\phi_m(\tau_t(X_1,\ldots,X_m)) = \rho(x) \tag{2.23}$$

Obviously from (2.22) - (2.23) we obtain (2.21) with

$$\sigma^2 = c\int_{-\infty}^{\infty}\rho(x)dx.$$

This behavior occurs in example (2.17): If $v(k) \sim \nu(k/m)$ for some function ν on $[0,1]$, then $c = f(\lambda)^2$ and $\rho(x) = (\nu * \nu(x)/\nu * \nu(0))^2$, cf. Welch (1967).

On the other hand it can also happen that the correlations remain stable, but the variance of ϕ_m increases.

$$\mathrm{Var}(\phi_m(X_1,\ldots,X_m)) \sim c_1 m \tag{2.24}$$

$$\lim_{m\to\infty}\sum_{t=-\infty}^{\infty}\mathrm{Corr}(\phi_m(X_1,\ldots,X_m),\phi_m(\tau_t(X_1,\ldots,X_m)) = c_2 \tag{2.25}$$

In this case $\sigma^2 = c_1c_2$. Such a behavior occurs in examples (2.14) and (2.15). For (2.14) this follows from a straightforward calculation. For (2.15) we replace the truncated innovations $X_m - \sum_{j=1}^{m-1} a_jX_{m-j}$ by the full innovations e_m. Then (2.24) follows from a simple calculation and (2.25) is clear, at least for martingale differences (e_t). See Berk (1974) for details.

Politis and Romano (1992) replace θ_n^m by

$$\theta_n^{m,L} = T^m(Q^{-1}\sum_{j=0}^{Q-1}\Delta(\tau_{jL}(X_1,\ldots,X_m))) \tag{2.26}$$

where $Q = (n-m)/L + 1$ and $L \sim am$ with $a \in (0,1]$.

Assuming that we have the analogue of (2.13):

$$\theta_n^{m,L} = Q^{-1}\sum_{j=0}^{Q-1}\phi_m(\tau_{jL}(X_1,\ldots,X_m)) + R_{m,n,L}$$

with negligible remainder term, the efficiency loss of $\theta_n^{m,L}$ depends very much on whether (2.22) – (2.23) or (2.24) – (2.25) hold. In the first case the efficiency is

$$\int_{-\infty}^{\infty}\rho(x)dx/(a\sum_{j=-\infty}^{\infty}\rho(ja))$$

which is close to one for a close to zero. On the other hand, if (2.24) – (2.25) holds, the efficiency is zero because $\theta_n^{m,L} - \theta^m$ is $O_p(mn^{-1/2})$ whereas $\theta_n^m - \theta^m$ is $O_p(m^{1/2}n^{-1/2})$. For the nonparametric spectral density estimation we saw that both types of behavior can occur for essentially the same estimator, simply by taking different versions. In general it is however not clear if there always exists a version such that (2.22) – (2.23) holds. For instance for the AR-spectrum (2.10) we are not able to construct such a version. So the modification (2.26) is not recommendable in all cases.

3 Bootstrapping the estimators of Section 2

We choose a block length ℓ and assume for simplicity that $b = n/\ell$ for (2.4) or $b = (n - m + 1)/\ell$ for (2.8) is an integer. Then we select the starting points of the blocks, $S_1, \ldots, S_b$, i.i.d. uniform on $\{0, 1, \ldots, n - \ell\}$ or on $\{0, 1, \ldots, n - m - \ell + 1\}$ respectively. The bootstrapped statistic is

$$\theta_n^* = T(n^{-1} \sum_{j=1}^{b} \sum_{t=0}^{\ell-1} \Delta(\tau_{S_j+t}\, X^{\text{ext}})) \tag{3.1}$$

in case (2.4) and

$$\theta_n^{m*} = T^m((n - m + 1)^{-1} \sum_{j=1}^{b} \sum_{t=0}^{\ell-1} \Delta(\tau_{S_j+t}(X_1, \ldots, X_m))) \tag{3.2}$$

in case (2.8). If the sample size is not a multiple of ℓ, we simply make the last block shorter. In (3.1) we do not take the variability of the mean in extending our sample into account. In order to do this, we can replace X^{ext} by $X^{*,\text{ext}}$ where

$$X_t^{*,\text{ext}} = \begin{cases} X_t & \text{if} \quad 1 \le t \le n \\ \overline{X}^* = n^{-1} \sum_{j=1}^{b} \sum_{t=1}^{\ell} X_{S_j+t} & \text{if} \quad t < 1 \text{ or } t > n. \end{cases}$$

We conjecture that the difference between using $X^{*,\text{ext}}$ and X^{ext} are usually negligible.

In the examples from the previous sections, this leads to the following bootstrapped statistics. For (2.6)

$$\theta_n^* = n^{-1} \sum_{j=1}^{b} \sum_{t=1}^{\ell} (X_{S_j+t} - \overline{X})[\lambda(X_{S_j+t} - \overline{X}) + 2 \sum_{k=1}^{n-S_j-t} (X_{S_j+t+k} - \overline{X}) \sin(k\lambda)/k]$$

For (2.7) we have

$$\theta_n^* = \text{argmin}_\alpha \sum_{j=1}^{b} \sum_{t=1}^{\ell} (X_{S_j+t} - \overline{X} - \sum_{k=1}^{S_j+t-1} \alpha^k (X_{S_j+t-k} - \overline{X}))^2. \tag{3.3}$$

For (2.8) we have

$$\begin{aligned} \theta_n^{m*} &= (n - m + 1)^{-1} \sum_{j=1}^{b} \sum_{t=1}^{\ell} (X_{S_j+t} - \overline{X})((X_{S_j+t} - \overline{X}) \\ &+ 2 \sum_{k=1}^{m-1} w(k, m)(X_{S_j+t+k} - \overline{X}) \cos(k\lambda)), \end{aligned} \tag{3.4}$$

and finally for (2.10) we have

$$\theta_n^{m*} = \sigma_{m-1}^{*2} |1 - \sum_{k=1}^{m-1} a_{k,m-1}^* \exp(i\lambda k)|^{-2} \tag{3.5}$$

where

$$(a_{1,m-1}^*, \ldots, a_{m-1,m-1}^*) = \text{argmin} \sum_{j=1}^{b} \sum_{t=1}^{\ell} (X_{S_j+t+m-1} - \overline{X} - \sum_{k=1}^{m-1} a_k (X_{S_j+t+m-1-k} - \overline{X}))^2$$

and σ^{*2}_{m-1} is the minimum of the function on the right hand side.

The procedure to prove consistency of the bootstrap is similar to the procedure used to show asymptotic normality for the statistic. With (2.4) we linearize T at F to obtain

$$\theta_n^* - \theta_n = n^{-1} \sum_{j=1}^{b} \sum_{t=0}^{\ell-1} \phi(\tau_{S_j+t} X) - n^{-1} \sum_{t=1}^{n} \phi(\tau_t X) + R_n^*. \tag{3.6}$$

With (2.8) we linearize T^m at F^m to obtain

$$\begin{aligned} \theta_n^{m*} - \theta_n^m &= (n-m+1)^{-1} \sum_{j=1}^{b} \sum_{t=0}^{\ell-1} \phi_m(\tau_{S_j+t}(X_1, \ldots, X_m)) \\ &- (n-m+1)^{-1} \sum_{t=1}^{n-m+1} \phi_m(\tau_t(X_1, \ldots, X_m)) + R_{m,n}^* \end{aligned} \tag{3.7}$$

For the remainder term one has to show that $R_n^* = o_p^*(n^{-1/2})$ and
$R_{m,n}^* = o_p^*((n/m)^{-1/2})$ respectively, in probability. This means that for any $\varepsilon, \delta > 0$

$$P[P^*[|R_n^*| > \varepsilon n^{-1/2}] > \delta] \longrightarrow 0$$

and

$$P[P^*[|R_{m,n}^*| > \varepsilon (m/n)^{1/2}] > \delta] \longrightarrow 0.$$

Finally one has to show that the linear part is asymptotically normal with the correct variance. For this one exploits that under P^* the linear part is an arithmetic mean of b i.i.d. variables since $S_1, \ldots, S_b$ are by definition i.i.d.. Details of these steps are given in the next two sections where we consider the examples (2.9) and (2.10).

Above we centered θ_n^* at θ_n and θ_n^m. Because of boundary effects this results in θ_n^* and θ_n having slightly different biases. In order to avoid this we could center θ_n^* at

$$\theta_n' = T((n-\ell+1)^{-1} \ell^{-1} \sum_{j=0}^{n-\ell} \sum_{t=0}^{\ell-1} \Delta(\tau_{j+t} X^{\mathrm{ext}}))$$

and

$$\theta_n^{m'} = T^m((n-m-\ell+2)^{-1} \ell^{-1} \sum_{j=0}^{n-m-\ell+1} \sum_{t=0}^{\ell-1} \Delta(\tau_{j+t}(X_1, \ldots, X_m)))$$

respectively. This means we center at the expectation of the bootstrap empirical. However this is relevant only when studying higher order approximations.

Finally let us briefly discuss what happens if we take two functionals T^m and U^m which coincide for stationary distributions. Then the bootstrap with U^m and T^m are different. Consider for instance

$$\begin{aligned} T^m &= E[\phi_m(X_1, \ldots, X_m)] \\ U^m &= E[\overline{\phi}_m(X_1, \ldots, X_m)] \end{aligned}$$

where $\overline{\phi}_m(X_1,\ldots,X_m)=\phi_m(X_1,\ldots,X_m)+g(X_1,\ldots,X_{m-1})-g(X_2,\ldots,X_m)$. Calculating the bootstrap (3.2) with T^m and U^m leads to the difference

$$(n-m+1)^{-1}\sum_{j=1}^{b}\{g(X_{S_j+1},\ldots,X_{S_j+m-1})-g(X_{S_j+\ell+1},\ldots,X_{S_j+\ell+m-1})\}$$

For fixed m this is typically of order $O_p^*(\ell^{-1/2}n^{-1/2})$ and thus asymptotically negligible. The difference in the bootstrap variances based on T^m and U^m is then of the order $O(\ell^{-1})$ which is the order of the bias of the bootstrap variance (Künsch, 1989, Theorem 3.2(i) and (3.15)). From the expression for the bias given in this Theorem it is also obvious that the bias is zero if $\overline{\phi}_m(X_1,\ldots,X_m)$ and $\overline{\phi}_m(X_j,\ldots,X_{j+m-1})$ are uncorrelated for $j\neq 1$. However to choose g such that $\overline{\phi}_m$ become uncorrelated is difficult and in any case requires the distribution F^∞ of (X_t) to be known (cf. Künsch 1984, Theorem 1.3). So it seems impossible to exploit this nonuniqueness to remove the bias of the bootstrap variance.

4 Proofs for the nonparametric estimator of the spectral density

Consider now the bootstrapped linearized estimator θ_n^{m*} as given in (3.7) with $R_{m,n}^*=0$:

$$\theta_n^{m*}=\theta^m+(n-m+1)^{-1}\sum_{j=1}^{b}\sum_{t=1}^{\ell}\phi_{S_j+t,m}, \tag{4.1}$$

where $\phi_{t,m}=\phi_m(\tau_{t-1}(X_1,\ldots,X_m)),E[\phi_{t,m}]=0$. In passing from (3.7) to (4.1) we have used (2.13) with $R_{m,n}=0$.

Let us consider first $\mathrm{Var}^*(\theta_n^{m*})$ as a variance estimator for $\mathrm{Var}(\theta_n^m)$, where $\theta_n^m=(n-m+1)^{-1}\sum_{t=1}^{n-m+1}\phi_{t,m}$.

By independence of (S_j):

$$\mathrm{Var}^*(\theta_n^{m*})=b^{-1}\mathrm{Var}^*(\ell^{-1}\sum_{t=1}^{\ell}\phi_{S_1+t,m})=(n-m+1)^{-1}A_1-b^{-1}A_2^2, \tag{4.2}$$

where

$$A_1=E^*[(\ell^{-1/2}\sum_{t=1}^{\ell}\phi_{S_j+t,m})^2]=(n-m-\ell+2)^{-1}\sum_{j=0}^{n-m-\ell+1}(\ell^{-1/2}\sum_{t=1}^{\ell}\phi_{j+t,m})^2 \tag{4.3}$$

and

$$A_2=E^*[\ell^{-1}\sum_{t=1}^{\ell}\phi_{S_1+t,m}]. \tag{4.4}$$

For the consistency of $\mathrm{Var}^*(\theta_n^{m*})$ we compute the expectation and variance of A_1 and A_2 respectively. Asymptotic normality of $(n/m)^{1/2}(\theta_n^{m*}-E^*[\theta_n^{m*}])$ will be shown by bounding the third moment of $\ell^{-1/2}\sum_{t=1}^{\ell}\phi_{S_1+t,m}$ and applying the Berry-Esséen result.

The conditions required and techniques used depend on whether we are in situation (2.22) - (2.23) or (2.24) - (2.25). We first treat the former which is easier. Let $\alpha(n)$ denote the strong mixing coefficients of (X_t) introduced by Rosenblatt (1956).

Theorem 4.1 *Let θ_n^{m*} be given by (4.1). Assume that $(X_t)_{t\in\mathbb{Z}}$ is stationary, strong-mixing with*

$$\sum_{i=0}^{\infty}(i+1)\alpha(i)^{\delta/(4+\delta)} < \infty \quad \text{and} \quad E|\phi_{1,m}|^{4+2\delta} = O(1), \delta > 0,$$

and (2.23) holds for any $|x| \leq 1$. Moreover assume that $\ell \to \infty, m \to \infty, \ell = o(n)$, $m = o(\ell), \ell m^{-1} n^{-1/2} = o(1)$. Then:

1) $nm^{-1}(\mathit{Var}^*(\theta_n^{m*}) - \mathit{Var}(\theta_n^m)) = o_p(1)$,
2) $\sup_{x\in\mathbf{R}} |P^*[(n/m)^{1/2}(\theta_n^{m*} - E^*[\theta_n^{m*}]) \leq x] - P[(n/m)^{1/2}(\theta_n^m - E[\theta_n^m]) \leq x]| = o_p(1).$

Proof: For part 1) we use the representation (4.2) for $\mathrm{Var}^*(\theta_n^{m*})$ and consider first A_1. We decompose

$$A_1 = (n-m-\ell+2)^{-1} \sum_{j=0}^{n-m-\ell+1} Z_j + E|\ell^{-1/2}\sum_{t=1}^{\ell}\phi_{t,m}|^2,$$

where $Z_j = Y_j - E[Y_j], Y_j = \left(\ell^{-1/2}\sum_{t=j+1}^{j+\ell}\phi_{t,m}\right)^2$.

Then:

$$\begin{aligned} &E[m^{-1}A_1] - (n-m+1)m^{-1}\mathrm{Var}(\theta_n^m) = \\ &m^{-1}\sum_{h=-\ell+1}^{\ell-1}\mathrm{Cov}(\phi_{0,m},\phi_{h,m})(\frac{|h|}{n-m+1} - \frac{|h|}{\ell}) - \\ &2m^{-1}\sum_{h=\ell}^{n-m}\mathrm{Cov}(\phi_{0,m},\phi_{h,m})(1-\frac{|h|}{n-m+1}) = o(1) \end{aligned} \tag{4.5}$$

by (2.23). Furthermore

$$\mathrm{Var}(m^{-1}A_1) = (n-m-\ell+2)^{-2}m^{-2}\sum_{j_1,j_2=0}^{n-m-\ell+1} E[Z_{j_1}Z_{j_2}]. \tag{4.6}$$

To estimate $E[Z_{j_1}Z_{j_2}]$ we look at two different cases.

i) If $|j_1 - j_2| > \ell + m - 1$ we can bound the covariance with the mixing coefficients (cf. Yokoyama (1980)).

$$|(n-m-\ell+2)^{-2}m^{-2}\sum_{\substack{j_1,j_2=0\\|j_1-j_2|>\ell+m-1}}^{n-m-\ell+1} E[Z_{j_1}Z_{j_2}]| \leq \text{const.}(n-m-\ell+2)^{-1}m^{-2}\|Z_1\|_{2+\delta}^2, \tag{4.7}$$

where $\|Z\|_p = E[|Z|^p]^{1/p}$ denotes the L_p-norm.

ii) If $|j_1 - j_2| \leq \ell + m - 1$ then by the Cauchy-Schwarz inequality:

$$|(n-m-\ell+2)^{-2}m^{-2} \sum_{\substack{j_1,j_2=0 \\ |j_1-j_2|\leq \ell+m-1}}^{n-m-\ell+1} E[Z_{j_1}Z_{j_2}]| \leq \|Z_1\|_2^2(\ell+m)(n-m-\ell+2)^{-1}m^{-2}. \tag{4.8}$$

By Minkowski's inequality:

$$\|Z_1\|_{2+\delta}^2 \leq \left(\ell\|\phi_{1,m}\|_{4+2\delta}^2\right)^2 = O(\ell^2) \tag{4.9}$$

On the other hand:

$$\|Z_1\|_2^2 = Var(Z_1) \leq E[(\ell^{-1/2}\sum_{t=1}^{\ell}\phi_{t,m})^4] = O(m^2) \tag{4.10}$$

because the mixing coefficents $(\alpha_\phi(k))$ of $(\phi_{t,m})$ satisfy

$$\alpha_\phi(k) \leq \begin{cases} 1 & \text{for } |k| \leq m-1 \\ \alpha(k-m+1) & \text{otherwise.} \end{cases}$$

Hence (4.10) follows from explicit bounds for the fourth moment of a sum of strong-mixing random variables (cf. Yokoyama (1980)). Using (4.6) - (4.10) we arrive at:

$$\mathrm{Var}(m^{-1}A_1) = O(\ell^2 m^{-2} n^{-1}) + O(\ell n^{-1}). \tag{4.11}$$

By combining (4.5) and (4.11):

$$m^{-1}A_1 - (n-m+1)m^{-1}\mathrm{Var}(\theta_n^m) = o_p(1). \tag{4.12}$$

To complete the proof of part 1) we have to analyze

$$A_2 = (n-m-\ell+2)^{-1}\sum_{t=1}^{n-m+1} c(n,t)\phi_{t,m},$$

where $c(n,t) = \min(t\ell^{-1}, 1, (n-m-t+2)\ell^{-1})$.

By a straightforward calculation we get:

$$A_2 = O_p((m/n)^{1/2}). \tag{4.13}$$

For part 2) we consider

$$(n-m+1)^{1/2}m^{-1/2}(\theta_n^{m*} - E^*[\theta_n^{m*}]) = b^{-1/2}\sum_{j=1}^{b}(U_n(S_j) - E^*[U_n(S_j)]),$$

where $U_n(S_j) = \ell^{-1/2}m^{-1/2}\sum_{t=1}^{\ell}\phi_{S_j+t,m}$.

By the Berry-Esséen Theorem and the independence of the (S_j) it suffices to show

$$\frac{E^*|U_n(S_1) - E^*[U_n(S_1)]|^3}{\mathrm{Var}^*(U_n(S_1))^{3/2}} = O_p(1)$$

From part 1) we know

$$Var^*(U_n(S_1)) \longrightarrow \sigma_\infty^2 > 0 \quad \text{in probability}.$$

Thus we have to show:

$$E^*|U_n(S_1) - E^*[U_n(S_1)]|^3 = O_p(1). \tag{4.14}$$

Denote by $\|.\|_{*p}$ the L_p-norm with respect to P^*. Trivially

$$\|U_n(S_1) - E^*[U_n(S_1)]\|_{*3} \leq \|U_n(S_1)\|_{*4} + E^*[U_n(S_1)]. \tag{4.15}$$

Again using known results for fourth moments of sums of strong-mixing random variables (cf. Yokoyama (1980)) we arrive at

$$E[E^*|U_n(S_1)|^4] = O(1)$$

and so by Markov's inequality

$$E^*|U_n(S_1)|^4 = O_p(1). \tag{4.16}$$

On the other hand by using (4.13):

$$E^*[U_n(S_1)] = \ell^{1/2} m^{-1/2} A_2 = O_p(1). \tag{4.17}$$

By (4.15) - (4.17) we have proved (4.14) and thus part 2. □

Remarks: 1) Theorem 1 has been formulated so as to keep the moment and mixing conditions low while imposing more restrictions on the block size ℓ. But we can do without the condition $\ell m^{-1} n^{-1/2} = o(1)$ if we assume $\sum_{i=0}^\infty (i+1)^2 \alpha(i)^{\delta/(6+\delta)} < \infty$ and $E|\phi_{1,m}|^{6+\delta} = O(1), \delta > 0$. This can be seen by replacing the crude Minkowski inequality in (4.9) by the bound for the 6^{th} moment of the sum of strong-mixing variables (cf. Yokoyama (1980)).
2) By a more careful analysis of the error terms appearing in the proof of Theorem 1 we find that

$$\begin{aligned} nm^{-1}(\text{Var}^*(\theta_n^{m*}) - \text{Var}(\theta_n^m)) &= -m\ell^{-1} c \int_{-1}^{1} \rho(x)|x| dx (1 + o(1)) \\ &+ O_p(\max\{\ell m^{-1} n^{-1/2}, \ell^{1/2} n^{-1/2}\}). \end{aligned}$$

Aside from the $O_p(\ell m^{-1} n^{-1/2})$ term this is the same as the results of Künsch (1989) for fixed m. Under the conditions mentioned in Remark 1 above the $O_p(\ell m^{-1} n^{-1/2})$ term disappears. Note however that in general there is also a bias term $\theta^m - \theta$ which has to be taken into account when making inference about θ. For a discussion of this problem see Politis and Romano (1992, p. 2003).

As an example consider the tapered and averaged periodogram for a fixed $\lambda \neq 0, \pi$:

$$\theta_n^m = U^m(F_n^m) = (n - m + 1)^{-1} \sum_{t=0}^{n-m} \|v\|_2^{-2} \left| \sum_{k=1}^{m} X_{t+k} v(k) e^{i\lambda k} \right|^2, \tag{4.18}$$

where $\|v\|_2^2 = \sum_{k=1}^m v(k)^2$ (see (2.11) and (2.17)).

For simplicity we assume here that $E[X_t]$ is known and equal to zero. Asymptotically this makes no difference. The functional U^m is linear, so we can directly apply Theorem 1. Assuming that $v(k) \sim \nu(k/m)$ we have to verify

$$m^{-(4+2\delta)} E \left| \sum_{k=1}^{m} X_k v(k) e^{i\lambda k} \right|^{8+4\delta} = O(1). \tag{4.19}$$

In order to be able to deal with the coefficients $v(k)e^{i\lambda k}$, we need an even exponent, i.e. $\delta = 1/2$. Then (4.19) would follow under a suitable mixing condition and $E|X_1|^{10+\delta'} < \infty$ $(\delta' > 0)$ which is quite restrictive. Surprisingly we can get by with fewer moments by using a truncation technique.

Theorem 4.2 *Let θ_n^m be given by (4.18) with $v(k) \sim \nu(k/m)$, ν bounded. Assume that $(X_t)_{t\in\mathbb{Z}}$ is stationary, strong-mixing with*

$$\sum_{i=0}^{\infty} (i+1)^8 \alpha(i)^{\delta/(10+\delta)} < \infty \quad (0 < \delta < 1) \text{ and } E|X_1|^s < \infty \text{ for } s > 4.$$

Moreover assume that in addition to the conditions of Theorem 4.1 on ℓ and m, $mn^{-1/(s\gamma)} \longrightarrow \infty$, where $\gamma = \gamma(s,\delta) = 2(10+\delta)/(5(10+\delta-s))$.
Then the assertions 1) and 2) of Theorem 1 remain true.

Remark: The implications of the condition $mn^{-1/(s\gamma)} \to \infty$ can be read for $\delta \le 1/2$ from the following table:

s	4+	5	6	7
$1/(s\gamma) \le$	0.39	0.26	0.18	0.12

Proof: We truncate the variables X_t, apply Theorem 4.1 to the truncated variables and show that the effect of truncation is negligible.
Let

$$X_t' = \begin{cases} X_t & \text{if } |X_t| \le m^\gamma \\ \operatorname{sign}(X_t) m^\gamma & \text{if } |X_t| > m^\gamma. \end{cases}$$

Define

$$\phi_{t,m}' = \|v\|_2^{-2} \left| \sum_{k=1}^{m} X_{t+k-1}' v(k) e^{i\lambda k} \right|^2$$

(compare with (4.18)).

Then:

$$\begin{aligned} & P[\phi_{t,m}' \ne \phi_{t,m} \text{ for some } t \in \{1,\dots,n-m+1\}] \\ \le\ & P[X_t' \ne X_t \quad \text{for some } t \in \{1,\dots,n\}] \\ \le\ & \sum_{t=1}^{n} P[X_t' \ne X_t] \le nm^{-s\gamma} \to 0 \end{aligned} \tag{4.20}$$

since $mn^{-1/(s\gamma)} \to \infty$.

We show that the moment assumption for $\phi'_{t,m}$ in Theorem 1 holds, i.e., we show that

$$E|\phi'_{t,m}|^5 = O(1). \tag{4.21}$$

Define $X''_t = X'_t - E[X'_t]$. Applying and Hölder's inequality we get:

$$\begin{array}{lll} E[X'_t] & \leq & m^{-(s-1)\gamma}\|X_t\|_s, \\ E|X''_t|^r & \leq & \text{const. } m^{(r-s)\gamma}E|X_t|^s, \ r \geq s. \end{array}$$

By using these bounds we arrive at:

$$\begin{array}{lll} E^{1/10}|\phi'_{t,m}|^5 & \sim & \|m^{-1/2}\sum_{k=1}^m X'_k v(k)e^{i\lambda k}\|_{10} \\ & \leq & \|m^{-1/2}\sum_{k=1}^m X''_k v(k)e^{i\lambda k}\|_{10} + |m^{-1/2}E[X'_1]| = O(1), \end{array}$$

because it can be shown inductively as in the proof of Theorem 1 of Yokoyama (1980) that

$$\|m^{-1/2}\sum_{k=1}^{m} X''_k v(k)e^{i\lambda k}\|_{10} = O(1).$$

In order to carry Yokoyama's argument through we need $\|X''_1\|^r_{r+\delta} = O(m^{r^*-1})$ for $r \leq 5$, where $r^* = a$ for $r = 2a, 2a-1$. This explains the value for $\gamma = \gamma(s,\delta)$ in our theorem.

(4.20) says that we can replace $\phi_{t,m}$ by $\phi'_{t,m}$, (4.21) gives us the moment condition of Theorem 4.1. Finally we have to verify (2.23) for $\phi'_{t,m}$ and $|x| \leq 1$. This follows in a straightforward way from $E|X_1|^s < \infty$ for $s > 4$ and (4.20).

□

5 Proofs for the AR-spectral estimator

We consider in this section the AR-spectral estimator (2.10) and its bootstrap (3.5). As described in section 2, the terms ϕ_m in the linearization (2.13) behave as in (2.24) - (2.25). This makes the consistency proof of the bootstrap more difficult than with the behavior (2.22) - (2.23). Moment inequalities for mixing sequences as used in the previous section are not sufficient. In order to have more structure available, we consider here linear processes whose innovations are a martingale difference. Our setup is as follows:

$$X_t = \sum_{j=0}^{\infty} b_j \ e_{t-j} \tag{5.1}$$

with $b_0 = 1$ and

$$\sum_{j=0}^{\infty} |b_j| < \infty, \tag{5.2}$$

$$B(z) = \sum_{j=0}^{\infty} b_j z^j \neq 0 \text{ for all } |z| \leq 1. \tag{5.3}$$

Then by Wiener's theorem

$$\sum_{j=0}^{\infty} |a_j| < \infty \tag{5.4}$$

where a_j is the coefficient of z^j in the expansion of $A(z) = 1/B(z)$. Hence if (e_t) is stationary with $E|e_t| < \infty$, we also have

$$X_t = \sum_{j=1}^{\infty} a_j X_{t-j} + e_t. \tag{5.5}$$

Let $\mathcal{F}_t$ denote the σ-algebra generated by $e_s, s \leq t$. We make the following assumptions about (e_t) and $m = m(n)$:

$$E[e_t|\mathcal{F}_{t-1}] \equiv 0, \tag{5.6}$$

$$E[e_t^2|\mathcal{F}_{t-1}] \equiv \sigma_e^2 < \infty, \tag{5.7}$$

$$E[|e_t|^r|\mathcal{F}_{t-1}] \leq C_r < \infty \text{ a.s. for } r = 3, 4, \tag{5.8}$$

$$m = O(n^{1/3-\delta}) \text{ for some } \delta > 0, \tag{5.9}$$

$$\sum_{j=1}^{\infty} |a_{m+j}| = o(n^{-1/2}). \tag{5.10}$$

Under (5.1) - (5.3) and (5.6) - (5.10) Berk (1974) has shown that the remainder $R_{m,n}$ in (2.13) is asymptotically negligible and that for any fixed $\lambda \neq 0, \pi$

$$(n/m)^{1/2}(\theta_n^m - \theta^m) \xrightarrow{d} \mathcal{N}(0, 2f^2(\lambda)).$$

The next result thus implies consistency of the bootstrap.

Theorem 5.1 *Assume (5.1) - (5.3) and (5.6) - (5.10) and* $\ell = O(n^{2/3}), m = o(\ell)$ *and* $\ell m^2 = o(n)$. *Then for any fixed* $\lambda \neq 0, \pi$

$$(n/m)^{1/2}(\theta_n^{m*} - \theta_n^m) \xrightarrow{d^*} \mathcal{N}(0,\ 2f(\lambda)^2)$$

in probability.

Proof: We follow the strategy outlined in section 3. First we have to show that the remainder $R_{m,n}^*$ in (3.7) is negligible, i.e. $R_{m,n}^* = o_p^*((n/m)^{-1/2})$ in probability. The next step consists in replacing the truncated innovation $X_t - \sum_{j=1}^{m-1} a_j X_{t-j}$ by the full innovation e_t. For this we have to show that

$$(n - m + 1)^{-1} \sum_{j=1}^{b} \sum_{t=0}^{\ell-1} (\phi_m(\tau_{S_j+t}(X_1, \ldots, X_m)) - \psi_{S_j+t+1,m})$$

is $o_p^*((n/m)^{-1/2})$, where ϕ_m is defined in (2.15) and

$$\psi_{t,m} = -\gamma_m(\lambda)^T R(m-1)^{-1} \underline{X}_t^{t+m-2} e_{t+m-1}. \tag{5.11}$$

These two steps can be handled by long, but straightforward arguments along the lines of Theorem 2 and Theorem 6 of Berk (1974). At this point we use the additional assumptions (5.9) - (5.10) and $\ell = O(n^{2/3})$. In order to save space we do not give details here.

We now look more closely at the linearized statistic and show that

$$(n/m)^{1/2}(n-m+1)^{-1}\sum_{j=1}^{b}\sum_{t=1}^{\ell}(\psi_{S_j+t,m}-E^*[\psi_{S_j+t,m}])\xrightarrow{d^*}\mathcal{N}(0,\ 2f(\lambda)^2)$$

in probability. First we investigate

$$(n-m+1)m^{-1}\mathrm{Var}^*((n-m+1)^{-1}\sum_{j=1}^{b}\sum_{t=1}^{\ell}\psi_{S_j+t,m})=m^{-1}A_1-\ell m^{-1}A_2^2$$

where A_1 and A_2 are defined as in (4.3) and (4.4).

By assumptions (5.6), (5.7) and the definition of $\psi_{t,m}$

$$E[m^{-1}A_1]=m^{-1}\ell^{-1}E[(\sum_{t=1}^{\ell}\psi_{t,m})^2]=m^{-1}E[\psi_{0,m}^2]=m^{-1}\sigma_e^2\gamma_m(\lambda)^T R(m-1)^{-1}\gamma_m(\lambda).$$

So by Theorem 3 of Berk (1974),

$$E[m^{-1}A_1]\longrightarrow 2f(\lambda)^2. \tag{5.12}$$

(In Berk's notation γ_m contains the additional factor $m^{-1/2}$).

Moreover

$$\mathrm{Var}(m^{-1}A_1)=(n-m-\ell+2)^{-2}\ell^{-2}m^{-2}\times \tag{5.13}$$
$$\sum_{j_1,j_2=0}^{n-m-\ell+1}\ \sum_{t_1,t_2=j_1+1}^{j_1+\ell}\ \sum_{t_3,t_4=j_2+1}^{j_2+\ell}\mathrm{Cov}(\psi_{t_1,m}\psi_{t_2,m},\psi_{t_3,m}\psi_{t_4,m}).$$

We show now that

$$\mathrm{Var}(m^{-1}A_1)=o(1). \tag{5.14}$$

By (5.6) and the definition of $\psi_{t,m}$ it follows that

$$\mathrm{Cov}(\psi_{t_1,m}\psi_{t_2,m},\psi_{t_3,m}\psi_{t_4,m})=0 \text{ if } \max(t_1,t_2,t_3)<t_4. \tag{5.15}$$

Hence we have by simple combinatorial arguments

$$\mathrm{Var}(m^{-1}A_1)\le\ \text{const.}\ (n-m-\ell+2)^{-1}\ell^{-1}m^{-2}\times \tag{5.16}$$
$$\sum_{1\le t_1\le t_2\le\ell}\ \sum_{k=0}^{n}|\mathrm{Cov}\ (\psi_{t_1,m}\psi_{t_2,m},\psi_{t_2+k,m}^2)|.$$

The covariances on the right hand side of (5.16) are estimated in Lemma 5.1 below. Inserting these estimates we obtain

$$\begin{aligned}\mathrm{Var}(m^{-1}A_1)\le&\ \text{const.}\ (n-m-\ell+2)^{-1}\ell\sum_{k=1}^{n}\sum_{r=k-m}^{k-1}|b_r|\\&+\text{const.}\ (n-m-\ell+2)^{-1}(m\sum_{k=1}^{n}\sum_{r=k-m}^{k-1}|b_r|+\sum_{k=1}^{n}\sum_{r=k-m}^{\infty}|b_r|)\\&+\text{const.}\ (n-m-\ell+2)^{-1}\ell m^2.\end{aligned}\tag{5.17}$$

(The first term on the right-hand side occurs for $t_1 < t_2$ and $k > 0$, the second term for $t_1 = t_2$ and $k > 0$ and the third for $k = 0$). Hence

$$\mathrm{Var}(m^{-1}A_1) = O(\ell m/n) + O(m^2/n) + o(1) + O(\ell m^2/n)$$

and (5.14) follows from our assumptions about $(\ell, m) = (\ell(n), m(n))$.

Next we consider A_2. The expectation of A_2 is zero. By a simple calculation (see also (4.13)) we find $\mathrm{Var}(A_2) = O(m/n)$. So

$$A_2 = O_p((m/n)^{1/2}). \tag{5.18}$$

Therefore by (5.12), (5.14) and (5.18)

$$n/m\mathrm{Var}^*((n-m+1)^{-1}\sum_{j=1}^{b}\sum_{t=1}^{\ell}\psi_{S_j+t,m}) \longrightarrow 2f^2(\lambda).$$

in probability.

To complete the proof, we apply the Berry-Esséen Theorem. Using the same notation as in the proof of Theorem 4.1 we have to show

$$E^*[U_n(S_1)\ ] = O_p(1), \tag{5.19}$$

$$E^*[U_n(S_1)^4] = O_p(1). \tag{5.20}$$

Because $E^*[U_n(S_1)] = (\ell/m)^{1/2}A_2$, (5.19) follows from (5.18). To show (5.20) we bound $E[E^*[U_n(S_1)^4]]$ and apply Markov's inequality. But this last expression can be written in terms of fourth moments of $(\psi_{t,m})$. It can be handled with the same arguments as in the proof of (5.14).

□

It remains to prove the bounds for fourth moments of (ψ_t) used in (5.17).

Lemma 5.1 Assume (X_t) satisfies (5.1) - (5.3) and (5.6) - (5.8) and let $\psi_{t,m}$ be as defined in (5.11). Then for $k > 0$ and $j > 0$

1) $|\mathrm{Cov}(\psi_{t,m}\psi_{t+j,m}, \psi^2_{t+j+k,m})| \leq$ const. $m^2\sum_{r=k-m}^{k-1}|b_r|$,

2) $|\mathrm{Cov}(\psi^2_{t,m}, \psi^2_{t+k,m})| \leq$ const. $m^3\sum_{r=k-m}^{k-1}|b_r| +$ const. $m^2\sum_{r=k-m}^{\infty}|b_r|$,

3) $|\mathrm{Cov}(\psi_{t,m}\psi_{t+k,m}, \psi^2_{t+k,m})| \leq$ const. m^4,

4) $|\mathrm{Var}(\psi^2_{t,m})| \leq$ const. m^4.

Proof: We write $-\gamma_m(\lambda)^T R(m-1)^{-1} = (d_1, \ldots, d_{m-1})$. Since the sum of the moduli of the elements in any row of $R(m-1)^{-1}$is bounded uniformly in m and the row index (c.f. Hong-Zhi et al., (1982)), we have

$$|d_s| \leq \text{ const. } < \infty \text{ for all } s. \tag{5.21}$$

Using (5.1), we can write

$$\psi_{t,m} = \sum_{s=1}^{m-1} \sum_{r=0}^{\infty} d_s b_r \; e_{t+m-1-s-r} \; e_{t+m-1}.$$

Hence for $j \geq 0, k \geq 0$

$$\begin{aligned} &\mathrm{Cov}(\psi_{t,m}\psi_{t+j,m}, \psi^2_{t+j+k,m}) = \sum_{s_1,\ldots,s_4=1}^{m-1} \sum_{r_1,\ldots,r_4=0}^{\infty} d_{s_1} d_{s_2} d_{s_3} d_{s_4} b_{r_1} b_{r_2} b_{r_3} b_{r_4} \times \\ &(E[e_{t-s_1-r_1} e_t \; e_{t+j-s_2-r_2} \; e_{t+j} \; e_{t+j+k-s_3-r_3} \; e_{t+j+k-s_4-r_4} \; e^2_{t+j+k}] - \\ &E[e_{t-s_1-r_1} \; e_t \; e_{t+j-s_2-r_2} \; e_{t+j}] E[e_{t-s_3-r_3} \; e_{t-s_4-r_4} \; e^2_t]). \end{aligned} \tag{5.22}$$

In the fourth and eigth moments above at most four time indices can coincide, so by (5.8) each expectation is bounded by a constant. Hence the cases 3) and 4) follow immediately from (5.21). Moreover by (5.6) and (5.7) most of the moments in (5.22) are zero. To prove 1) and 2) we just have to look carefully for the non-zero terms.

Consider first the case $j > 0$ and $k > 0$. Then by (5.6) and (5.7) we have to consider

$$E[e_{t-s_1-r_1} \; e_t \; e_{t+j-s_2-r_2} \; e_{t+j} \; e_{t+j+k-s_3-r_3} \; e_{t+j+k-s_4-r_4}].$$

This does not vanish in the following 3 cases:

$$s_3 + r_3 = s_4 + r_4 = k; \tag{5.23}$$

$$s_3 + r_3 = k, \; s_2 + r_2 = j, \; s_4 + r_4 = j + k + s_1 + r_1; \tag{5.24}$$

$$s_3 + r_3 = k, \; s_2 + r_2 = j + s_1 + r_1, \; s_4 + r_4 = j + k. \tag{5.25}$$

Of course in (5.24) and (5.25) we can also exchange (s_3, r_3) with (s_4, r_4), but this gives just an additional factor of 2. Condition (5.23) means that $k - m \leq r_3, r_4 \leq k - 1$ and that s_3, s_4 are fixed when r_3 and r_4 are chosen. Hence summing the terms on the right-hand side of (5.22) for those indices satisfying (5.23) we obtain the bound

$$\text{const. } m^2 (\sum_{r=k-m}^{k-1} |b_r|)^2 \leq \text{ const. } m^2 \sum_{r=k-m}^{k-1} |b_r|.$$

Similarly summing over indices satisfying (5.24) gives the bound

$$\text{const. } m \sum_{r=k-m}^{k-1} |b_r| \sum_{r=j-m}^{j-1} |b_r| \sum_{r=j+k+1-m}^{\infty} |b_r| \leq \text{ const. } m \sum_{r=k-m}^{k-1} |b_r|.$$

Finally summing over indices satisfying (5.25) gives

$$\text{const. } m \sum_{r=k-m}^{k-1} |b_r| \sum_{r=j+1-m}^{\infty} |b_r| \sum_{r=j+k-m}^{j+k-1} |b_r| \leq \text{ const. } m \sum_{r=k-m}^{k-1} |b_r|.$$

Together this proves 1).

Finally for $j = 0$ and $k > 0$ we have to consider by (5.6) and (5.7)

$$E[e_{t-s_1-r_1} \; e_{t-s_2-r_2} \; e^2_t \; e_{t+k-s_3-r_3} \; e_{t+k-s_4-r_4}] - E[e_{t-s_1-r_1} \; e_{t-s_2-r_2}] E[e_{t-s_3-r_3} \; e_{t-s_4-r_4}].$$

This does not vanish in the following cases:

$$s_3 + r_3 = k,\ \ s_4 + r_4 \leq k; \tag{5.26}$$

$$s_3 + r_3 = k + s_1 + r_1,\ \ s_4 + r_4 = k + s_2 + r_2; \tag{5.27}$$

$$s_3 + r_3 = s_4 + r_4 = k + s_1 + r_1,\ \ s_2 + r_2 < s_1 + r_1. \tag{5.28}$$

Summation over the indices satisfying one of these conditions proves 2). □

We close this section with two remarks.

Remarks: 1) Validity of the bootstrap holds also if we look at the joint distribution of the AR-spectral estimator at a fixed number of frequencies. This follows easily with the Cramér-Wold device for proving multivariate central limit theorems. 2) If we assume the e_t's to be i.i.d., then additional terms on the right-hand side of (5.22) vanish and we can prove Theorem 5.1 without the additional condition $\ell m^2 = o(n)$. Details are left to the reader.

References:

Berk, K.N. (1974). *Consistent autoregressive spectral estimates.* Ann. Statist. **2**, 489-502.

Doob, J.L. (1953). *Stochastic Processes.* Wiley, New York.

Efron, B. (1979). *Bootstrap methods: Another look at the jackknife.* Ann. Statist. **7**, 1-26.

Hong-Zhi, A., Zhao-Guo, C. and Hannan, E.J. (1982). *Autocorrelation, autoregression and autoregressive approximation.* Ann. Statist. **10** , 926-936.

Künsch, H.R. (1984). *Infinitesimal robustness for autoregressive processes.* Ann. Statist. **12**, 843-863.

Künsch, H.R. (1989). *The jackknife and the bootstrap for general stationary observations.* Ann. Statist. **17**, 1217-1241.

Politis, D.N. and Romano, J.P. (1992). *A general resampling scheme for triangular arrays of α-mixing random variables.* Ann. Statist. **20**, 1985-2007.

Rosenblatt, M. (1956). *A central limit theorem and a strong mixing condition.* Proc. Natl. Acad. Sci. USA **42**, 43-47.

Welch, P.D. (1967). *The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms.* IEEE Trans. Audio Electroacoustics **15**, 70-73.

Yokoyama, R. (1980). *Moment bounds for stationary mixing sequences.* Z. Wahrsch. verw. Geb. **52**, 45-57.